

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2001, 10:01:43 ; Search time 2123.11 Seconds

(without alignments)  
16612.879 Million cell updates/sec

Title: US-09-215-035-1

Perfect score: 2138  
Sequence: 1 AGCAATCCGTCGCCGCC.....AAAAAAAAAAAAAAAAAAAA 2138

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenBank:\*

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_ph:\*  
7: gb\_pl:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_hum:\*  
32: em\_htgo\_hum:\*  
33: em\_htgo\_hum:\*  
34: em\_htg\_hum:\*  
35: em\_htg\_hum:\*  
36: em\_htg\_hum:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	100.0	2138	6	AR100763
2	2138	100.0	2138	6	AR119934
3	2114	98.9	2114	6	HS040434
4	1994	93.3	2129	6	E07943
5	1994	93.3	2129	6	E07943
6	1981.8	92.7	2093	9	BC003512
7	1975	92.4	2093	9	H00000
8	1961	91.7	2162	9	BC009272
9	980.8	45.9	1195	9	AF180951
10	910	42.6	2135	10	D86370
11	902.8	42.2	2099	10	D87351
12	307.6	14.4	37307	9	HS3357
13	307.6	14.4	154615	2	AC067772
14	307.6	14.4	256073	9	AE006464
15	296.6	13.9	155770	2	AC092381
16	278.2	13.0	155770	2	AC092381
17	243.2	11.4	185697	2	AC067814
18	219.8	10.3	396	6	AX093308
19	197	9.2	197	6	189984
20	117	5.3	209826	2	AC068497
21	75	3.5	75	6	189983
22	69	3.2	69	6	189982
23	54.2	2.5	2658	1	STM00000
24	53.8	2.5	188123	2	AC012004
25	53.2	2.5	2490	1	AB019394
26	53.2	2.5	23990	1	SC6E10
27	52.8	2.5	89948	9	HS756523
28	52.6	2.5	10732	6	E32986
29	52.4	2.5	2884	9	HSN800512
30	52.4	2.5	3103	9	HS10N
31	52.4	2.5	3137	9	BC000235
32	52.4	2.5	3179	9	HS002389
33	52.4	2.5	3407	6	176203
34	52.4	2.5	3407	6	176203
35	52	2.4	92458	9	AL553715
36	51.6	2.4	1736	9	BC009285
37	51.4	2.4	208936	2	AC010821
38	51.2	2.4	27210	9	HS422E10
39	51.2	2.4	278911	9	AE006465
40	50.6	2.4	298166	2	AC087563
41	50.4	2.4	114793	9	AF217796
42	50.2	2.3	2000	6	108490
43	50.2	2.3	41944	3	AC005929
44	50	2.3	1263	10	MH070658
45	50	2.3	1292	10	AF009520

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN
1	AR100763	Sequence 1 from patent US 6083502.	AR100763	AR100763.1	GI:12811561	Unknown.	1 (bases 1 to 2138)	Pastan, I. and Chang, K.	Mesothelium antigen and methods and kits for targeting it	Patent: US 6083502-A 1 04-JUL-2000; Location/Qualifiers	385 a	722 c 661 g 370 t
						Unknown.				1..2138		
						Unclassified.				/organism="unknown"		
						14-FEB-2001						

Query Match	100.0%;	Score 2138;	DB 6;	Length 2138;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2138;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	AGAAATTCGGGTGGCCGGGACACTCCGCTGGCTGTATACGCGGGACAGAGACTACCGGT	60
Db	1	AGGAATTCGGGTGGCCGGGACACTCCGCTGGCTGTATACGCGGGACAGAGACTACCGGT	60
QY	61	GGACCCACAGGTGCTCCCTCCCTGGGATCTACACAGACCATGGCTTTCAGACGGCTTCAC	120
Db	61	GGACCCACAGGTGCTCCCTCCCTGGGATCTACACAGACCATGGCTTTCAGACGGCTTCAC	120
QY	121	CCCTGTGTGTCCTCTGTGGGGACCGCCCTGGCAGCTCTCTGTTCCTGTCTTTCAGCTTCGA	180
Db	121	CCCTGTGTGTCCTCTGTGGGGACCGCCCTGGCAGCTCTCTGTCTTTCAGCTTCGA	180
QY	181	TGGGTGCATCCCGGAGAGACCTTGGCTGGAGACACAGGAGACGAGTCTCCCTGGG	240
Db	181	TGGGTGCATCCCGGAGAGACCTTGGCTGGAGACACAGGAGACGAGTCTCCCGG	240
QY	241	GGAGTCCCTGACACACCCCCCAATTTCCAGCTCTCCCTTCGCCAACCTCTTGGCTTC	300
Db	241	GGAGTCCCTGACACACCCCCCAATTTCCAGCTCTCCCTTCGCCAACCTCTTGGCTTC	300
QY	301	CCGTTGCGGAGGTGTCCGGGCTGAGACAGGACGCTCCGGAGTGGCTGGCTGGGCTTG	360
Db	301	CCGTTGCGGAGGTGTCCGGGCTGAGACAGGACGCTCCGGAGTGGCTGGCTGGGCTTG	360
QY	361	GCACAGAGAATGTCAAGCTCTCACAGAGACAGTCCGCTGTCTGGCTTACCGGCTCTCT	420
Db	361	GCACAGAGAATGTCAAGCTCTCACAGAGACAGTCCGCTGTCTGGCTTACCGGCTCTCT	420
QY	421	GAGCCCCCAGAGACCTGAGACGCCCTTCCATTGGACTCTGCTATTCCTCAACCCAGAT	480
Db	421	GAGCCCCCAGAGACCTGAGAGGCCCTTCCATTGGACTCTGCTATTCCTCAACCCAGAT	480
QY	481	GCCTTCTCGGGGCCCCAGGCGCTGACCCGTTTCTTCCCGGATCACAGAAAGCCAAATMG	540
Db	481	GCCTTCTCGGGGCCCCAGGCGCTGACCCGTTTCTTCCCGGATCACAGAAAGCCAAATMG	540
QY	541	GACCTGCTCCCGAGGGGGGCTCCCGAGACGACAGCGGCTCTGCTGCGGCTCTTGACCTGAC	600
Db	541	GACCTGCTCCCGAGGGGGGCTCCCGAGACGACAGCGGCTCTGCTGCGGCTCTTGACCTGAC	600
QY	601	TGGGGTGTGGGGGGGTCTCTCTGCTGAGCAGAGCTGATGTGCGGGCTCTTGGAGGCTGCT	660
Db	601	TGGGGTGTGGGGGGGTCTCTCTGCTGAGCAGAGCTGATGTGCGGGCTCTTGGAGGCTGCT	660
QY	661	TGCACACTGCTGGGCGCTTGTGGCGGATGCGCGGAGTCTGTATACCCGGCTGGTG	720
Db	661	TGCACACTGCTGGGCGCTTGTGGCGGATGCGCGGAGTCTGTGTATACCCGGCTGGTG	720
QY	721	AGCTGCCCCGAGACCCCTGAGCACAGACACAGAGAGGACAGCGGCGGCTCTGACAGGC	780
Db	721	AGCTGCCCCGAGACCCCTGAGCACAGACACAGAGAGGACAGCGGCGGCTCTGACAGGC	780
QY	781	GGGGGACACCCCTACAGGGCCCCCGGTGGACATGCTGTCTTCCACATGGACGCTTGGG	840
Db	781	GGGGGACACCCCTACAGGGCCCCCGGTGGACATGCTGTCTTCCACATGGACGCTTGGG	840
QY	841	GGCCTGCTGCCCTGTGTGGGCGACGCCATCATTCGAGCATCCCGAAGCATTCGTGGC	900
Db	841	GGCCTGCTGCCCTGTGTGGGCGACGCCATCATTCGAGCATCCCGAAGCATTCGTGGC	900
QY	901	GCGTGTGGGCAAGCTCTCTCGGGACCCATCTGGCGGACCTGAAAGCAGCATCTCTC	960
Db	901	GCGTGTGGGCAAGCTCTCTCGGGACCCATCTGGCGGACCTGAAAGCAGCATCTCTC	960
QY	961	CGCGCGGGGTTCGGGGGGGAAGTGGAGAGACAGCTGTCTTTCAGGGAAGAAGGCCCGC	1020
Db	961	CGCGCGGGGTTCGGGGGGGAAGTGGAGAGACAGCTGTCTTTCAGGGAAGAAGGCCCGC	1020

QY	1021	GATATAGACGAGACCTCATCTTCTCTACAGAAATGGAGACTGGAAGCCCTGGCTGGATCGC	1080
DB	1021	GAGATAGACGAGACCTCATCTTCTCTACAAAAATGGGAGCTGGAAAGCTGGTGGATCGC	1080
QY	1081	GCCCTGCTGGGCGACCCAGATGAGACCGGCTGAAGCGCCATCCCTTCACCTACGAGACAGTG	1140
DB	1081	GCCCTGCTGGGCGACCCAGATGAGAGCGGCTGAAGCGCCATCCCTTCACCTACGAGAGAGTG	1140
QY	1141	GACGTCTAAAGCATAAACTGATGAGCTCTACCCCAAGGTTTACCCCGAGTCTGTGATC	1200
DB	1141	GACGTCTAAAGCATAAACTGATGATGAGCTCTACCCCAAGGTTTACCCCGAGTCTGTGATC	1200
QY	1201	CAGGACCTGGGGCTACCTCTTCTCTCAGATGAGCGCCCTGAGGACATTTCCGAACTGGAAATGTG	1260
DB	1201	CAGGACCTGGGGCTACCTCTTCTCTCAAAATGAGCGCTGAGAGCATTTCCGAAGTGGAAATGTG	1260
QY	1261	ACGTCCCTGGAGACCTCGAAGGCTTTGCTTGAAGTGTGACAAAGGGCGACGAAATGAGTCT	1320
DB	1261	ACGTCCCTGGAGACCTCGAAGGCTTTGCTTGAAGTGTGACAAAGGGCGACGAAATGAGTCT	1320
QY	1321	CAGGCTCTCGGGGGCCCCCTCCACAGAGTGCGCACCTGTATCGACCGCTTTGTGAAGGA	1380
DB	1321	CAGGCTCTCGGGGGCCCCCTCCACAGAGTGGCGCACCTGTATCGACCGCTTTGTGAAGGA	1380
QY	1381	AGGGGCGAGCTAAGCAAAAGACCCCTAAGACCCCTGACCGCTTCTACCTGGGTACTTG	1440
DB	1381	AGGGGCGAGCTAAGCAAAAGACCCCTAAGACCCCTGACCGCTTCTACCTGGGTACTTG	1440
QY	1441	TGCTTCCTCACCCCCGAGAGAGCTGAGTCCGTCGCCCCAGCAGCATCTGGGCGGTCAAG	1500
DB	1441	TGCTTCCTCACCCCCGAGAGAGCTGAGTCCGTCGCCCCAGCAGCATCTGGGCGGTCAAG	1500
QY	1501	CCCCAGACCTGGAACAGTGTGACCCCAAGGCACTGGAAGTCTCTATCCCAAGGCCGC	1560
DB	1501	CCCCAGACCTGGAACAGTGTGACCCCAAGGCACTGGAAGTCTCTATCCCAAGGCCGC	1560
QY	1561	CTTGCTTTCAGAAACATGAACGGGTCCGAATACTTGTAATAATCCACTCTTCTCGGT	1620
DB	1561	CTTGCTTTCAGAAACATGAACGGGTCCGAATACTTGTAATAATCCACTCTTCTCGGT	1620
QY	1621	GAGGCCCCCAGGAGATTGTAAGCGGCCTACGACACAAATGTGAGCATGAGACTTGAGC	1680
DB	1621	GAGGCCCCCAGGAGATTGTAAGCGGCCTACGACACAAATGTGAGCATGAGACTTGAGC	1680
QY	1681	ACGTTTCATGAAGCTGCGGAGCGATGGGGTCTGCCCTTGACTGTGGTGAAGTGAAGAA	1740
DB	1681	ACGTTTCATGAAGCTGCGGAGCGATGGGGTCTGCCCTTGACTGTGGTGAAGTGAAGAA	1740
QY	1741	CTTCTGGGAGCCCAACGATGAGAGGGCTCTGAAGCGGAGAGAGCCGACCGCCGCTGGGAC	1800
DB	1741	CTTCTGGGAGCCCAACGATGAGAGGGCTCTGAAGCGGAGAGAGCCGACCGCCGCTGGGAC	1800
QY	1801	TGAGATCTACGGCAGCGGCGAGGAGACACTGAGACAGCTGGGGCTGGGGCTCAAGGGCGGC	1860
DB	1801	TGAGATCTACGGCAGCGGCGAGGAGACACTGAGACAGCTGGGGCTGGGGCTCAAGGGCGGC	1860
QY	1861	ATCCCAACGGCTACCTGATCTAGAACCTGAGGTGAAGAGAACCTCTCGGGAGCGCC	1920
DB	1861	ATCCCAACGGCTACCTGATCTAGAACCTGAGGTGAAGAGAACCTCTCGGGAGCGCC	1920
QY	1921	TGCTCTCTAGAACCTGGACCTGTTCTACCGCTCTGAGCACTGCTCTTCAAGCTCCACCTG	1980
DB	1921	TGCTCTCTAGAACCTGGACCTGTTCTACCGCTCTGAGCACTGCTCTTCAAGCTCCACCTG	1980
QY	1981	GCTGAGGGGCCCCACCTCTCTTGTGTGGCCCCAGCCCTGCTGGGGATCCCGGCTGGCAAG	2040
DB	1981	GCTGAGGGGCCCCACCTCTCTTGTGTGGCCCCAGCCCTGCTGGGGATCCCGGCTGGCAAG	2040
QY	2041	AGCAGGCGACGGGTATCCCGCTTCCACCCCAAGAGAACTCGCGCTCAGTAAACGGGAACA	2100
DB	2041	AGCAGGCGACGGGTATCCCGCTTCCACCCCAAGAGAACTCGCGCTCAGTAAACGGGAACA	2100
QY	2101	TGCCCCCTGGCAGCAAAAAAAAAAAAAAAAAAAAAAAAA 2138	

Db 2101 TGCCCCCTGCAGACAAAAA 2138

RESULT 2

ARL19934 2138 bp DNA PAT 16-MAY-2001

LOCUS ARL19934

DEFINITION Sequence 1 from patent US 6153430.

ACCESSION ARL19934

VERSION ARL19934.1 GI:14102633

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2138)

AUTHORS Pastan,I. and Chang,K.

TITLE Nucleic acid encoding mesothelin, a differentiation antigen present on mesothelioma, mesotheliomas and ovarian cancers

JOURNAL Patent: US 6153430-A 1 28-NOV-2000;

FEATURES

location/Qualifiers

1..2138

BASE COUNT 385 a 722 c 661 g 370 t

ORIGIN

Query Match 100.0%; Score 2138; DB 6; Length 2138;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAATTCGGGTGGCCGGCCACATCCCTGCTGCTGAGCGCGGACAGAGAGCTAACCGT 60

Db 1 AGGAATTCGGGTGGCCGGCCACATCCCTGCTGCTGAGCGCGGACAGAGAGCTAACCGT 60

QY 61 GGACCCAGGAGGCTCCCTCCCTGGGATCTACAGAGACATGCGCTTGCAACGCGTGCAC 120

Db 61 GGACCCAGGAGGCTCCCTCCCTGGGATCTACAGAGACATGCGCTTGCAACGCGTGCAC 120

QY 121 CCCGTTGGTCTGTGGGGACCGCCCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGGA 180

Db 121 CCCGTTGGTCTGTGGGGACCGCCCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGGA 180

QY 121 CCCGTTGGTCTGTGGGGACCGCCCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGGA 180

Db 121 CCCGTTGGTCTGTGGGGACCGCCCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGGA 180

QY 181 TGGGTGATCCCGCGAGAGACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240

Db 181 TGGGTGATCCCGCGAGAGACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240

QY 241 GGAATTCCTGAAACCCCGCCCAATATTTCCAGCTCTCCCTCCCGCAACTCTTGGCTTC 300

Db 241 GGAATTCCTGAAACCCCGCCCAATATTTCCAGCTCTCCCTCCCGCAACTCTTGGCTTC 300

QY 241 GGAATTCCTGAAACCCCGCCCAATATTTCCAGCTCTCCCTCCCGCAACTCTTGGCTTC 300

Db 241 GGAATTCCTGAAACCCCGCCCAATATTTCCAGCTCTCCCTCCCGCAACTCTTGGCTTC 300

QY 301 CCGGTGCGAGAGTGTCCGGCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

Db 301 CCGGTGCGAGAGTGTCCGGCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

QY 301 CCGGTGCGAGAGTGTCCGGCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

Db 301 CCGGTGCGAGAGTGTCCGGCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360

QY 361 GCACAGAAAGATGTCAAGCTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

Db 361 GCACAGAAAGATGTCAAGCTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

QY 421 GAGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480

Db 421 GAGCCCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 480

QY 481 GCGTTCTCGGGGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 540

Db 481 GCGTTCTCGGGGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 540

QY 541 GACCTGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 600

Db 541 GACCTGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 600

QY 601 TGGGGTGTGGGGGGGCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 660

Db 601 TGGGGTGTGGGGGGGCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 660

QY 661 TGCGACCTGCGGGGGGCTTTGTGGCGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAT 720

Db 661 TGCGACCTGCGGGGGGCTTTGTGGCGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAT 720

QY 721 AGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 780

Db 721 AGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 780

QY 781 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 840

Db 781 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 840

QY 841 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 900

Db 841 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 900

QY 901 GCGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 960

Db 901 GCGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 960

QY 961 GCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1020

Db 961 GCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1020

QY 1021 GAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1080

Db 1021 GAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1080

QY 1081 GCGCTGCTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140

Db 1081 GCGCTGCTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140

QY 1141 GAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1200

Db 1141 GAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1200

QY 1201 CAGACAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1260

Db 1201 CAGACAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1260

QY 1261 ACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1320

Db 1261 ACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1320

QY 1321 CAGGCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1380

Db 1321 CAGGCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1380

QY 1381 AGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1440

Db 1381 AGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1440

QY 1441 TGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1500

Db 1441 TGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1500

QY 1501 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1560

Db 1501 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1560

QY 1561 CTTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1620

Db 1561 CTTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1620

QY 1621 GGGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1680

Db 1621 GGGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1680

QY 1681 ACCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1740

Db 1681 ACCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1740





D	601	TGGGGTGTGGGGGGGGTCTCTGCTGACAGAGCGCTAMTGTGGGGCTCTTGGAGGGCTGGCT	660
Q	661	TGCGACCTGCCCTGGGGCGCTTGTGTGGCCGAGTGGCGCCGAATGTCTACCCGGCGTGGT	720
D	661	TGCGACCTGCCCTGGGGCGCTTGTGTGGCCGAGTGGCGCCGAAGTGTCTCTACCCGGCGTGGT	720
Q	721	AGCTGGCCCCGGAGACCCCTTGGACCGAGACACAGAGAGAGAGCCAGGGGGCGCTCTGCAGGGC	780
D	721	AGCTGGCCCCGGAGACCCCTTGGAGACCGAGACACAGAGAGAGAGCCAGGGGGCGCTCTGCAGGGC	780
Q	781	GGGGGAGACCCCCCTACAGGAGCCCCCGCTGAGATGATGTCTTCCACATATGAGATGTCTGGG	840
D	781	GGGGGAGACCCCCCTACAGGAGCCCCCGCTGAGATGATGTCTTCCACATATGAGATGTCTGGG	840
Q	841	GGCCCTGCTGGCCGCTGCTGGGGCAGCCCCATCATCCGAGCATCCCGACAGGCGATGTGGCC	900
D	841	GGCCCTGCTGGCCGCTGCTGGGGCAGCCCCATCATCCGAGCATCCCGACAGGCGATGTGGCC	900
Q	901	GCGTGGCGGCAAGCTCTCTCTGGGAGCCCATCTTGCGGGCAGCTGAAGSAGCATCTCTC	960
D	901	GCGTGGCGGCAAGCTCTCTCTGGGAGCCCATCTTGCGGGCAGCTGAAGSAGCATCTCTC	960
Q	961	CGGGCGGGGTTCCGGGGGGGAGAGAGACAGCTGTCTCTAGGAGAAAGAGCCCGC	1020
D	961	CGGGCGGGGTTCCGGGGGGGAGAGAGACAGCTGTCTCTAGGAGAAAGAGCCCGC	1020
Q	1021	GAGATAGACAGAGAGCTCATCTCTCAAGAGATGGGAGCTGGAGAGCTGGTGATAGC	1080
D	1021	GAGATAGACAGAGAGCTCATCTCTCAAGAGAGATGGGAGCTGGAGAGCTGGTGATAGC	1080
Q	1081	GCCTGCTGGCCACCCAGATGAGACCGCGTGAACGCCATCCCTTACTACGACAGCTG	1140
D	1081	GCCTGCTGGCCACCCAGATGAGACCGCGTGAACGCCATCCCTTACTACGACAGCTG	1140
Q	1141	GAGCTCTTAAGCATTAACGTGATGAGCTCTACCCACAAGGTATCCCGAGTGTGATC	1200
D	1141	GAGCTCTTAAGCATTAACGTGATGAGCTCTACCCACAAGGTATCCCGAGTGTGATC	1200
Q	1201	CAGCAGCTGGGCTACCTTCTCTCAAGATGAGCCCTGAGAGCATTTCCGAAGTGAATGTG	1260
D	1201	CAGCAGCTGGGCTACCTTCTCTCAAGATGAGCCCTGAGAGCATTTCCGAAGTGAATGTG	1260
Q	1261	ACGTCCCTGGAGACCTCTGAAGGCTTGTGTAAGTGTGACAAAGGCGACGAATGAGTCT	1320
D	1261	ACGTCCCTGGAGACCTCTGAAGGCTTGTGTAAGTGTGACAAAGGCGACGAATGAGTCT	1320
Q	1321	CAGGCTCTGGGGGGGCCCCCTCCACAGGGTGGCCACCTGATGACAGCGTTTGTGAAGGGA	1380
D	1321	CAGGCTCTGGGGGGGCCCCCTCCACAGGGTGGCCACCTGATGACAGCGTTTGTGAAGGGA	1380
Q	1381	AGGGGCGAGCTAGACAAAGACACCTAGACACCTGACCGCTTCTACTCTGGGTACTCTG	1440
D	1381	AGGGGCGAGCTAGACAAAGACACCTAGACACCTGACCGCTTCTACTCTGGGTACTCTG	1440
Q	1441	TGCTCCCTCAGCCCCGAGAGCTGAGCTCCGTGCCCCCGACGAGCATGTGGGGCGTACG	1500
D	1441	TGCTCCCTCAGCCCCGAGAGCTGAGCTCCGTGCCCCCGACGAGCATGTGGGGCGTACG	1500
Q	1501	CCCCGAGACCTGGAGACGTGTGACCCCAAGGAGCTGGAGATCTCTATATCCCAAGGCCCG	1560
D	1501	CCCCGAGACCTGGAGACGTGTGACCCCAAGGAGCTGGAGATCTCTATATCCCAAGGCCCG	1560
Q	1561	CTTCTCTTCCAGAAATGAAGCGGGTCCGAATCTTGTGAAGATCCATCTCTTCTGGGT	1620
D	1561	CTTCTCTTCCAGAAATGAAGCGGGTCCGAATCTTGTGAAGATCCATCTCTTCTGGGT	1620
Q	1621	GGGGCCCCCAGGAGATTTGAAGGCGCTTACGTACAGACAGATGTGAGCATGTGAGCTGGCC	1680
D	1621	GGGGCCCCCAGGAGATTTGAAGGCGCTTACGTACAGACAGATGTGAGCATGTGAGCTGGCC	1680
Q	1681	ACGTTCATGAAGCTGGCGAGAGATGGGGTGTGCGCTTACGTGTGGCTGTGAGGTGTGAGAA	1740
D	1681	ACGTTCATGAAGCTGGCGAGAGATGGGGTGTGCGCTTACGTGTGGCTGTGAGGTGTGAGAA	1740

OY	1741	CTTCTGGGAGACCCCACTGTGAAGGGCCTTGAAGCGGAGAGACCGGCACCACGCCGCTGGGGAC	1800
Db	1741	CTTTGGGAGACCCCACTGTGAAGGGCCTTGAAGCGGAGAGACCGGCACCACGCCGCTGGGGAC	1800
OY	1801	TGGATTCCTACGGCACGGGAGAGACACTTGAGACACGCTGGGGGCTGAGGCTACAGGGCGGC	1860
Db	1801	TGGATTCCTACGGCACGGGAGAGACACTTGAGACACGCTGGGGGCTGAGGCTACAGGGCGGC	1860
OY	1861	ATOCACAACGGCTACCTGGTGTCTTAGACCTTCAAGCGTGAAGAGAACCTTCTGGGGAGCGCC	1920
Db	1861	ATOCACAACGGCTACCTGGTGTCTTAGACCTTCAAGCGTGAAGAGAACCTTCTGGGGAGCGCC	1920
OY	1921	TGGCTTCCTAAGACCTTGACACGCTTCTTGACACGCTTCTTGACACGCTTCTTGACACGCTT	1980
Db	1921	TGGCTTCCTAAGACCTTGACACGCTTCTTGACACGCTTCTTGACACGCTTCTTGACACGCTT	1980
OY	1981	GCGTAAAGGCCCCCATCTCCCTTGCTGGGCCCCAGCCCTTGCTGGGGATCCCAGGCTGAGCAGG	2040
Db	1981	GCGTAAAGGCCCCCATCTCCCTTGCTGGGCCCCAGCCCTTGCTGGGGATCCCAGGCTGAGCAGG	2040
OY	2041	AGCAGCAGCAGGGGTGATCCCGCTTCCACACCCCAAGAGACACTGCGGCTCACTAAAGGGAACA	2100
Db	2041	AGCAGCAGCAGGGGTGATCCCGCTTCCACACCCCAAGAGACACTGCGGCTCACTAAAGGGAACA	2100
OY	2101	TGCCCCCTGTCAGAC	2114
Db	2101	TGCCCCCTGTCAGAC	2114
RESULT	4		
E07943			
LOCUS	E07943	2129 bp	RNA
DEFINITION		cDNA encoding a polypeptide having megakaryocyte potentiating activity.	PAT 29-SEP-1997
ACCESSION	E07943		
VERSION	E07943.1	GI:2176075	
KEYWORDS	JP 1994225767-A/1.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Yamauchi,M., Kojima,T., Ooeda,M. and Hattori,A.		
JOURNAL	GENE CODING MAGKARYOCYTE AMPLIFIER		
COMMENT	PATENT: JP 1994225767-A 1 15-AUG-1994; CHUGAI PHARMACEUT CO LTD OS Homo sapiens (human) PN JP 1994225767-A/1 PD 15-OCT-1994 PF 23-OCT-1992 JP 1993288617 PR 23-OCT-1992 JP 92P 236153, 11-NOV-1992 JP 92P 301387, PR PI YAMAUCHI MARE, KOJIMA TETSUO, OEODA MASAYOSHI, PI HATTORI ARIHITO PC C12N15/16, C12N1/21, C12N5/10, C12P21/02//A61K37/02, C07K13/00, PC C1201/68, PC (C12N1/21, C12R1:19), (C12N5/10, C12R1:91), (C12P21/02, C12R1:19), PC (C12P21/02, PC (C12R1:91); CC strandedness: Double; CC topology: Linear; CC hypothetical: No; CC anti-sense: No; FH Location/Qualifiers FT source 1..2129 FT /organism='Homo sapiens' FT /cell_line='HPC-Y5' FT /clone='pkp027' FT 1..96 FT 5'UTR FT CDS FT 97..1965 FT /product='polypeptide having megakaryocyte FT		

FT 3'UTR  
FT misc-feature  
FT  
FT  
Location/Qualifiers  
1. 2129  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 389 a 716 c 657 g 367 t  
ORIGIN

Query Match 93.3%; Score 1994; DB 6; Length 2129;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 2075; Conservative 0; Mismatches 35; Indels 30; Gaps 3;

potentiating  
activity  
1966..2129  
1873  
/note="The nucleotide at position 1873 of cDNA  
in other  
clone pKPO21 is A"

5 ATTCGGTGGCGGACCACTCCGTCGTCTGTGACGGCGGACAGAGACCTACCGGTGAC 64  
2 AATTCGGACAGGACCACTCCGTCGTCTGTGACGGCGGACAGAGACCTACCGGTGAC 61  
65 CCAGGAGCCCTCCCTCCCTGGGATCTACAGACCATGGCTTG-CAAGGCTGACCCC 123  
62 CCAGGAGCCCTCCCTCCCTGGGATCTACAGACCATGGCTTGCAAGGCTGACCCC 121  
124 TGTGTGTCCTGTGGGACCGCC-----TGAGAGCTCTCTGTCCTGCTTCAAGCTCG 178  
122 TGTGTGGGTCCTGTGGGACCGCCCGCCCTGGGACCTCTCTGCTCTTCAAGCTCG 181  
179 GATGGGTGCATCCCGAGAGACCTGTGGTGGAGAGACAGGAGGAGTCTGCCCGCTGG 238  
182 GATGGGTGCACCTCCAGAGACCTGTGGTGGAGAGAGGAGGAGGAGGAGGAGGAGG 241  
239 GGGAGTCTGACAAACCCCATTAACATTTCCAGCTCTCCCTCCGCACTCTGCTGCT 298  
242 ACGAGTCTGTGGACCCACCACTTAACATTTCCAGCTCTCCCTCCGCACTCTGCTGCT 301  
299 TCCGCTGTGGAGAGTGTCCGCTGTGACAGAGAGAGCTGCGGAGCTGTGGCT 358  
302 TCCGCTGTGGAGAGTGTCCGCTGTGACAGAGAGAGCTGCGGAGCTGTGGCT 361  
359 TGGCAGAGAAATGTCAAGCTCTCAACAGAGAGAGTGGCTGTGGCTCAACGGCTCT 418  
362 TGGCAGAGAAATGTCAAGCTCTCAACAGAGAGAGTGGCTGTGGCTCAACGGCTCT 421  
419 CTGAGCCCGCCGAGAGCTGTGAGAGCCCTCCCATGAGAGCTCTCTAATTCCTCAAC 478  
422 CTGAGCCCGCCGAGAGCTGTGAGAGCCCTCCCATGAGAGCTCTCTAATTCCTCAAC 481  
479 ATGCGTCTGGGGGCCCCAGAGCTGTGACAGAGCTCTCTCCGCAATGAGAGGCAATG 538  
482 ATGCGTCTGGGGGCCCCAGAGCTGTGACAGAGCTCTCTCCGCAATGAGAGGCAATG 541  
539 TGGCAGAGAAATGTCAAGCTCTCAACAGAGAGAGTGGCTGTGGCTCAACGGCTCT 598  
542 TGGCAGAGAAATGTCAAGCTCTCAACAGAGAGAGTGGCTGTGGCTCAACGGCTCT 601  
599 GCTGGGGTGTGGGGGCTCTCTGTGAGAGAGCTGATGTGCGGGCTCTGGGAGGCTGG 658  
602 GCTGGGGTGTGGGGGCTCTCTGTGAGAGAGCTGATGTGCGGGCTCTGGGAGGCTGG 661  
659 CTTCGAGACTGCTGGGGGCTTTGTGGCGAGTGGGCGAAGTGTCTGACCCGGGCTGG 718  
662 CTTCGAGACTGCTGGGGGCTTTGTGGCGAGTGGGCGAAGTGTCTGACCCGGGCTGG 721  
719 TGAGCTCCCGGAGACCCCTGAGACAGAGACGAGAGAGGAGGAGGAGGAGGAGGAGG 778  
722 TGAGCTCCCGGAGACCCCTGAGACAGAGACGAGAGAGGAGGAGGAGGAGGAGGAGG 781  
779 GCGGGGAGACCCCTACGGGCCCCCGTGCAGATGTGTCTGTCCAGATGAGGAGCTTGC 838

Db 782 GCGGGGAGACCCCTACAGCCGCCCGCTGCAGATGTGTCTCCAGATGAGAGCTTGC 841  
Oy 839 GGGGCTCTCTCCGCTGTGGGCAAGCCCATATCCGAGATCCCGAGGAGCATCGG 898  
Db 842 GGGGCTCTCTCCGCTGTGGGCAAGCCCATATCCGAGATCCCGAGGAGCATCGG 901  
Oy 899 CCGGCTGGGCAACGCTCTCTGGGAGCCCATCTTGGGCGAGCTGAGAGGAGCATTC 958  
Db 902 CCGGCTGGGCAACGCTCTCTGGGAGCCCATCTTGGGCGAGCTGAGAGGAGCATTC 961  
Oy 959 TCCGGGCGGGGCTTCCGGGAGAGTGGAGAGAGAGCTGTCTTCAAGAGAGAGGCC 1018  
Db 962 TCCGGGCGGGGCTTCCGGGAGAGTGGAGAGAGAGCTGTCTTCAAGAGAGAGGCC 1021  
Oy 1019 GCGAGATGAGAGAGAGCTCTCTTCAAGAGAGTGGAGAGAGCTGTCTTCAAGAGAG 1078  
Db 1022 GCGAGATGAGAGAGAGCTCTCTTCAAGAGAGTGGAGAGAGCTGTCTTCAAGAGAG 1081  
Oy 1079 GGGGCTCTGTGGGACCCAGATGGACCGGCTGAGAGGAGGAGGAGGAGGAGGAGG 1138  
Db 1082 GGGGCTCTGTGGGACCCAGATGGAGCGGCTGAGAGGAGGAGGAGGAGGAGGAGG 1141  
Oy 1139 TGGAGCTCTTAAGCATTAAGTGTGATGATGATGATGATGATGATGATGATGATG 1198  
Db 1142 TGGAGCTCTTAAGCATTAAGTGTGATGATGATGATGATGATGATGATGATGATG 1201  
Oy 1199 TCCAGACCTGGGCTACCTCTTCCATGATGAGGAGGAGGAGGAGGAGGAGGAGG 1258  
Db 1202 TCCAGACCTGGGCTACCTCTTCCATGATGAGGAGGAGGAGGAGGAGGAGGAGG 1261  
Oy 1259 TGGAGCTCTTGAAGCCTGAGAGGCTTCTTGAAGTGGAGAGAGGAGGAGGAGGAGG 1318  
Db 1262 TGGAGCTCTTGAAGCCTGAGAGGCTTCTTGAAGTGGAGAGAGGAGGAGGAGG 1321  
Oy 1319 CTGAGAGCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1378  
Db 1322 CT-----CAGGTGGCAGGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1387  
Oy 1379 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1438  
Db 1388 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1447  
Oy 1439 TGTGCTCTTCAAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1498  
Db 1448 TGTGCTCTTCAAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1477  
Oy 1499 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1558  
Db 1478 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1537  
Oy 1559 GCTTGTCTTCAAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1618  
Db 1538 GCTTGTCTTCAAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1597  
Oy 1619 GTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1678  
Db 1598 GTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1657  
Oy 1679 CCAGCTTCAAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1738  
Db 1658 CCAGCTTCAAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1717  
Oy 1739 AACTTGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1798  
Db 1718 AACTTGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1777  
Oy 1799 AACTGATCTAAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1858  
Db 1778 AACTGATCTAAGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1837  
Oy 1859 GCATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1918  
Db 1838 GCATCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1897

QY	1919	CTGCTCTCTAGACCTGGACCTGTTCTTCACGCTGCTGGCACTGCTCCAGCTCCACCC	1978
QY	1919	CTGCTCTCTAGACCTGGACCTGTTCTTCACGCTGCTGGCACTGCTCCAGCTCCACCC <td>1978</td>	1978
Db	1898	CCTGCTCTCTAGACCTGGACCTGTTCTTCACGCTGCTGGCACTGCTCCAGCTCCACCC	1957
QY	1979	TGGCCTGAGAGGCCACACTCCCTTGTGTGGCCCCAGGCCCTGCTGGGATCCCTCCCTGGCCA	2038
Db	1958	TGGCCTGAGAGGCCACACTCCCTTGTGTGGCCCCAGGCCCTGCTGGGATCCCTCCCTGGCCA	2017
QY	2039	GGAGCAGCAGCAGGGGTGATCCCGCTTCACCCCAAGAACTCGCTCACTAAAGGGAA	2098
Db	2018	GGAGCAGCAGCAGGGGTGATCCCGCTTCACCCCAAGAACTCGCTCACTAAAGGGAA	2077
QY	2099	CATGCCCCCTGCAGACAAAACCAAAAAAAAAAAAAAAAAAAAAA 2138	
Db	2078	CATGCCCCCTGCAGACAACTAATAAAAAAAAAAAAAAAAAAAAAA 2117	
RESULT	5		
LOCUS	189985	2129 bp	DNA
DEFINITION	Sequence 34 from patent US 5723318.	PAT	10-AUG-1998
ACCESSION	189985		
VERSION	189985.1	GI:3409925	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 2129)		
AUTHORS	Yamauchi, N., Kojima, T., Oh-eda, M. and Hattori, K.		
TITLE	DNA coding for megakaryocyte potentiator		
JOURNAL	Patent: US 5723318-A 34 03-MAR-1998;		
FEATURES	Location/Qualifiers		
source	1..2129		
BASE COUNT	389 a 716 c 657 g 367 t		
ORIGIN			
Query Match	93.3%; Score 1994; DB 6; Length 2129;		
Best Local Similarity	97.0%; Pred. No. 0;		
Matches 2075; Conservative	0; Mismatches 35; Indels 30; Gaps		
QY	5	ATTCCGGTGGCCGGCCACTCCCGCTGCTGTGAGCGCGGACAGAGAGCTACCGGTGAC	64
Db	2	AATTCGGCAGCAGAGCCACTCCCGCTGCTGTGAGCGCGGACAGAGAGCTACCGGTGAC	61
QY	65	CCAGGTCGCTCCCTCCCGTGGATCTACACAGACATGGCCCTTG-CAACGGCTCGACCCC	123
Db	62	CCAGGTCGCTCCCTCCCGTGGATCTACACAGACATGGCCCTTGCAACGGCTCGACCCC	121
QY	124	TGTTGGTCTCTGTGGAGGACGGCC-----TGGCAGCTCTCTTCTGCTCTTCAAGCTCG	178
Db	122	TGTTGGGTCCTGTGGAGAACCCCGCCCTCGGAGCTCTCTTCTGCTCTTCAAGCTCG	181
QY	179	GATGGGTGATCCCGCGAGAGACCTTGCTGTGAGAGACAGAGACGATGTGCCCCCTCG	238
Db	182	GATGGGTGATCCCGCGAGAGACCTTGCTGTGAGAGACAGAGAGGCTGTGACACCTCG	241
QY	239	GGGAGTCTGACACACCCCAATTAATTTCCAGGCTCTCCCTCGGCAACATCTTGGCT	298
Db	242	ACGGAGTCTTGGCCAAACCACTAATTAATTTCCAGGCTCTCCCTCGGCAACATCTTGGCT	301
QY	299	TCCCGTGTGGAGAGTGTCCGGCTGAGCAGCAGACGATGTCCGGAGCTGAGCTGTGAGCT	358
Db	302	TCCCGTGTGGAGAGTGTCCGGCTGAGCAGCAGACGATGTCCGGAGCTGAGCTGTGAGCT	361
QY	359	TGGCAGACAGAAATGTCAAGCTTCAACAGAGACAGCTGGCTGTCTGCTCAACGGCTCT	418
Db	362	TGGCAGACAGAAATGTCAAGCTTCAACAGAGACAGCTGGCTGTCTGCTCAACGGCTCT	421
QY	419	CTGAGCCCCCGAGAACTGTGAGACGCCCTTCCATTGGACTGTGTCTATTCGAACCCAG	478

Db	4422	CTGAGCCCCCGAGAGACTGTGAGCCCTCCATTGGACGTGCTATCTCAACCCAG	481
Qy	419	ATGCGTTCTGGGGCCCCAGGCTGCACCCGTTTCTTCTCCGCATACAGAAAGCCATG	538
Db	482	ATGGTTCTCGGGGCCCCAGGCCGTGACCCGTTTCTTCTCCGCATACAGAAAGCCATG	541
Qy	539	TGACACTGCTCCCGAGGGGGCTCCGAGAGAGGGGTGCTGCGCTGCGCTTGCGCT	598
Db	542	TGGACCTGCTCCGAGGGGGGGCTCCGAGAGAGGGGTGCTGCGCTGCGCTTGCGCT	601
Qy	559	GCTGGGGTGTCCGGGGTCTCTGCTGACCGAGCTATGTGGGGCTCTGGAGGCTTG	658
Db	602	GCTGGGGTGTCCGGGGTCTCTGCTGACCGAGCTATGTGGGGCTCTGGAGGCTTG	661
Qy	659	CTTGGGACCTGCTCGGGGCGGCTTGTGGCCGAGTGGGCCGGAAGTGGCTATCCCGGCTG	718
Db	662	CTTGGGACCTGCTCGGGGCGGCTTGTGGCCGAGTGGGCCGGAAGTGGCTATCCCGGCTG	721
Qy	719	TGAGCTGCCCGGAGCCCTTGACAGAGACCCAGAGAGGAGGAGCGGCTCTGACAG	778
Db	722	TGAGCTGCCCGGAGCCCTTGACAGAGACCCAGAGAGGAGGAGCGGCTCTGACAG	781
Qy	779	GCGGGGAGACCCCTTACGGGCCCCCGTGCACATGTCGTGCTGCACAGATGAGACGCTGCG	838
Db	782	GCGGGGAGACCCCTTACGGGCCCCCGTGCACATGTCGTGCTGCACAGATGAGACGCTGCG	841
Qy	839	GGGGCTGCTGCTCCGCTGTGGGCGAGCCCATCTCCGACAGCATCCGCGAGGGCATCTGG	898
Db	842	GGGGCTGCTGCTCCGCTGTGGGCGAGCCCATCTCCGACAGCATCCGCGAGGGCATCTGG	901
Qy	899	CCGGTGGCGGACAGCGTCCCTGCGACCATCTGCGGGGACGCTGGAACGGACCATCC	958
Db	902	CCGGTGGCGGACAGCGTCCCTGCGACCATCTGCGGGGACGCTGGAACGGACCATCC	961
Qy	959	TCCGGCCCGGTTCTCCGCGGGGAAGTGGAGAGAGACGCTGCTCTTCAGGCAAGAGGCC	1018
Db	962	TCCGGCCCGGTTCTCCGCGGGGAAGTGGAGAGAGACGCTGCTCTTCAGGCAAGAGGCC	1021
Qy	1019	GCGAGATGAGACGAGACGCTCATCTTCTACAAAGAAATGGGAGCTGGAAAGCTGCGTGGATG	1078
Db	1022	GCGAGATGAGACGAGACGCTCATCTTCTACAAAGAAATGGGAGCTGGAAAGCTGCGTGGATG	1081
Qy	1079	CGGCGCTGCTGGCACCACTAGATGAGACGCGCTGGAACCCATCCCTTCACCTACGAGACAGC	1138
Db	1082	CGGCGCTGCTGGCACCACTAGATGAGACGCGCGCTGGAACCCATCCCTTCACCTACGAGACAGC	1141
Qy	1139	TGAGACTCTTAAAGCATPAACTGATGACTTACCCACAGAGTTACCCGAGTCTGTGA	1198
Db	1142	TGAGACTCTTAAAGCATPAACTGATGAGCTTACCCACAGAGTTACCCGAGTCTGTGA	1201
Qy	1199	TCCAGCACTGGGGTACTCTTCTCTCAAGATGAGCCCTGGAGCATTTGGCAAGTGGAAATG	1258
Db	1202	TCCAGCACTGGGGTACTCTTCTCTCAAGATGAGCCCTGGAGCATTTGGCAAGTGGAAATG	1261
Qy	1259	TGACGTTCCCTGAGAGACCTGAAAGGCTTGTGAAGTGCACAAAGGGCACAGAAATGAGTTC	1318
Db	1262	TGACGTTCCCTGAGAGACCTGAAAGGCTTGTGAAGTGCACAAAGGGCACAGAAATGAGTTC	1321
Qy	1319	CTCAGAGCTCTCGGGGGGCCCTTCCCAAGTGGGCCACCTGATGCACCGCTTTGTGAAG	1378
Db	1322	CT-----CAGTGGGCACCCCTGATGCACCGCTTTGTGAAG	1381
Qy	1379	GAAAGGGGACCTGAGACAAGAGACCCCTTAACACCCCTGACACCGGCTTCTACCTTGGGTAC	1438
Db	1388	GAAAGGGGACCTGAGACAAGAGACCCCTTAACACCCCTGACACCGGCTTCTACCTTGGGTAC	1441
Qy	1439	TGTGCTCCTCAGCCCGAGAGAGCTGAGCTCTGTCGCCCCCAGACAGCATCTGGGCGGTCA	1498
Db	1448	TGTGCTCCTCAGCCCGAGAGAGCTGAGCTCTGTCGCCCCCAGACAGCATCTGGGCGGTCA	1501
Qy	1499	GGCCCCAGAGACTGGACAAGTGTGACCCAGAGAGAGTGGACAGTCTTATCCCAAGGGCC	1556
Db	1478	GGCCCCAGAGACTGGACAAGTGTGACCCAGAGAGAGTGGACAGTCTTATCCCAAGGGCC	1559

QY	1559	GCCTTGCTTTCCAGAAACATGAACGGGCTCCGAATACTTGCTGAAGATCCAGTCCCTCCGG	1618
Db	1538	GCCTTGCTTTCCAGAAACATGAACGGGCTCCGAATACTTGCTGAAGATCCAGTCCCTCCGG	1597
QY	1619	GTGGGGCCCCCAGGAGGATTTGAAGGGCCCTAGTCACAGAAATGTGAGCATGTGACTTGG	1678
Db	1598	GTGGGGCCCCCAGGAGGATTTGAAGGGCCCTAGTCACAGAAATGTGAGCATGTGACTTGG	1657
QY	1679	CCAGCTTCATGAAAGCTGGGGAGAGCATGTGGGTGCTGCCCTTACCTGTGGCTAGAGTGCAGA	1738
Db	1658	CCAGCTTCATGAAAGCTGGGGAGAGCATGTGGGTGCTGCCCTTACCTGTGGCTAGAGTGCAGA	1717
QY	1739	AACTCTTGGGAGACCCACGACTGGAGGGCTCTGAAGGGCGGAGACGGACCGCCCGGTCCAGG	1798
Db	1718	AACTCTTGGGAGACCCACGACTGGAGGGCTCTGAAGGGCGGAGAGACGGACCGCCCGGTCCAGG	1777
QY	1799	ACTGATCTCTACGAGCGAGCGGACGAGCATCTTGACACGCTTGGGGCTGAGAGGGCG	1858
Db	1778	ACTGATCTCTACGAGCGAGCGGAGCGAGCATCTTGACACGCTTGGGGCTGAGAGGGCG	1837
QY	1859	GCATCCCAACAGCGCTACCTGTGCTTCAAGACCTCAGGTGTCAGAGACCTCTCCGGGAGCG	1918
Db	1838	GCATCCCAACAGCGCTACCTGTGCTTCAAGACCTCAGGTGTCAGAGAGCCCTCTCCGGGAGCG	1897
QY	1919	CTGTGCTCTCTAGAGACCTTGAGCTGTTCTTCACCGTCCCTGGACACTGCTCTTACCTTCACCC	1978
Db	1898	CTGTGCTCTCTAGAGACCTTGAGCTGTTCTTCACCGTCTTGACACTGCTCTTACCTTCACCC	1957
QY	1979	TGGCTTGAGGGCCCCACATCCCTTGTGCTGGCCCCAGCCCTGTGGGGATCCCGCTTGAGCA	2038
Db	1958	TGGCTTGAGGGCCCCACATCCCTTGTGCTGGCCCCAGCCCTGTGGGGATCCCGCTTGAGCA	2017
QY	2039	GGAGCAGGACAGGGGTGATCCCGCTTCCACCGCAAGAGAACTCGGGCTCAGTAAACGGGAA	2098
Db	2018	GGAGCAGGACAGGGGTGATCCCGCTTCCACCGCAAGAGAACTCGGGCTCAGTAAACGGGAA	2077
QY	2099	CATGCCCCCTCGACAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2138
Db	2078	CATGCCCCCTCGACAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2117
RESULT	6		
BC003512			
LOCUS	BC003512	2433 bp	mRNA
DEFINITION	Homo sapiens, mesothelium, clone MGC:10686 IMAGE:3611296, mRNA, complete cds.		
ACCESSION	BC003512		
VERSION	BC003512.1	GI:13097581	
KEYWORDS	MGC.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2433)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submissions		
JOURNAL	Submitted (20-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgaps-rt@mail.nih.gov">cgaps-rt@mail.nih.gov</a> Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada <a href="mailto:info@cgsc.bc.ca">info@cgsc.bc.ca</a> Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Buterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Rand Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo		

Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeary, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranda Tsai, Natsija van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

clone distribution: MEC clone distribution information can be found through the I.M.A.G.E. Consortium/INL at: <http://Image.lnl.gov>  
Series: IRAL plate: 13 Row: a Column: 21  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 1129078.

FEATURES  
source

Location/Qualifiers
1. .2433

CDS

```
/note="Vector: p01B7"
418. :2283
```

BASE COUNT  
ORIGIN . .

431 a	823 c	755 g	424 t
-------	-------	-------	-------

Query Match	92.7%;	Score 1981.8;	DB 9;	Length 2433;
Best Local Similarity	97.2%;	Pred. No. 0;		
Matches: 2071;	Conservative	0;	Mismatches 27;	Indels 33; Caps 4;

QY 12 TGCCGGCCACTCCCGTCTGCTGTGACGCCTGGACAGAGAGCTACCGGTGACCCACGGT 71

Db. 330 TGGCCGGCACTCCCGTCTGCTGTGACGCCCGACAGAGCTACCGGTGACCCACGGT 38

72 GCCTCCCTCCCTGGGATCTACACAGACCATTGGCCTTG-CAACGGCTTCGACCCCTGTTGGT 13

Db 390 GCCTCCCTCCCTGGGATCTACACAGACCATGGCTTGCACAGGCTTCGACCCCTGTGGG 44

Q7. 131 CCTGTGGGACCGCC---TGGACGCTTCCTGTCTCTTCAGCCCTGGATGGGT 18

Db 450 GTCCTGTGGGACCCCGCCCTGCTGTCTGCTCTTCAGCCCTCGGATGGGT 50

Qy 186 GCATCCCCGGAGACCTGGCTGGAGAGACAGGAGCAGTCTGCCCCCTGGGGGAGT 24

Db 510 GCAGCCCTCGAGGACCTGGCTGGAGAGACAGG--GCAGGCTGCACCCCTGGACGGAGT 56

246 CCTGACACCCCCCATTAACATTTCAGGCTCTCCCCCTGGCCAACTCTTGGCTTCCCGTG 30

Db 567 CCTGGCCAAACCTACATTTCCAGCCCTCCCCCTGCCAACACTCTTGCGTTCCCGTG 62

306 TGGGAGGTGTCCGGCTGAGCACGGAGGCTGTCCGGAGCTGGCTGTGGCCTTGGCAC 36

b2 / TCCGAGCGTGTCCGGCTGAGCACGGAGCGTGTCGGAGCTGCGTTGGCAC 68

388 GAAGAAATGTCGAAGCTCTCCACACAGCAGCTGCGCTGCTGGGCTACCGGGGCTCTGAGCC 42

08/ 0A56A5A191C4A6C1C1C4A6C4A6C4C1B0C31G1C1B0C1C4C0C0C1C1C1B0A6C / 4

QY 426 CCCCCGAGACCTGAGACGCCCTCCATTGGACCTGCTGCTATTCTTCAACCCAGATGCGTT 485  
 |||||  
 Db 747 CCCCCGAGACCTGAGACGCCCTCCATTGGACCTGCTGCTATTCTTCAACCCAGATGCGTT 806  
 |||||  
 QY 486 CTGCGGGGCCCCAGGACCTGACACCGCTTCTCTCCGCAATCAGAAAGCCAAATGTGACCT 545  
 |||||  
 Db 807 CTCGGGGGCCCCAGGACCTGACACCGCTTCTCTCCGCAATCAGAAAGCCAAATGTGACCT 866  
 |||||  
 QY 546 GCTCCCAAGGGGGGGGGTCCCGAGCCAGACACCGGCTGCTGCTGCGGCTGTGGGCTGTGGG 605  
 |||||  
 Db 867 GCTCCCAAGGGGGGGGGTCCCGAGCCAGACACCGGCTGCTGCTGCGGCTGTGGGCTGTGGG 926  
 |||||  
 QY 606 TGTGCGGGGGGCTCTGCTGAGACGAGGCTGATGTCGGGCTGTGGGAGGCTGTGGGTTGGA 665  
 |||||  
 Db 927 TGTGCGGGGGGCTCTGCTGAGACGAGGCTGATGTCGGGCTGTGGGAGGCTGTGGGTTGGA 986  
 |||||  
 QY 666 CTTGCTTGGGGGCTTTTGTGGCCGAGTGGGCGGAAGTGTCTTACCCGGGCTGTGAGCTG 725  
 |||||  
 Db 987 CTTGCTTGGGGGCTTTTGTGGCCGAGTGGGCGGAAGTGTCTTACCCGGGCTGTGAGCTG 1046  
 |||||  
 QY 726 CCGGGGACCCCTGGACAGAGACAGAGAGGAGGAGGCGGCGGCTGTGACGGGCGGGGG 785  
 |||||  
 Db 1047 CCGGGGACCCCTGGACAGAGACAGAGAGGAGGAGGCGGCGGCTGTGACGGGCGGGGG 1106  
 |||||  
 QY 786 ACCCCCTACAGGCCCCCGCTGACATGGTCTGTCTCCAGATGAGACGCTGTGGGGGCT 845  
 |||||  
 Db 1107 ACCCCCTACAGGCCCCCGCTGACATGGTCTGTCTCCAGATGAGACGCTGTGGGGGCT 1166  
 |||||  
 QY 846 GCTGCGGCTGTGGGGCCAGCCCATCATCTCCGACATCCCGAGGCGCATCTGTGGCGGCTG 905  
 |||||  
 Db 1167 GCTGCGGCTGTGGGGCCAGCCCATCATCTCCGACATCCCGAGGCGCATCTGTGGCGGCTG 1226  
 |||||  
 QY 906 GCGGCAACGCTCTCTGCGGAGCCCATCTGCGGAGGCTGAGGAGGAGCATCTCTCCGCGC 965  
 |||||  
 Db 1227 GCGGCAACGCTCTCTGCGGAGCCCATCTGCGGAGGCTGAGGAGGAGCATCTCTCCGCGC 1286  
 |||||  
 QY 966 GCGGCTTCCGGGGGGAAGTGAAGAACAGCGCTGTCTTCAAGCAAGAGGCGCGGAGAT 1025  
 |||||  
 Db 1287 GCGGCTTCCGGGGGGAAGTGAAGAACAGCGCTGTCTTCAAGCAAGAGGCGCGGAGAT 1346  
 |||||  
 QY 1026 AGAGGAGAGCTCATCTTCTTCAAGAAAGTGGAGCTGGAGACCTGCTGTGATGCGGCGCT 1085  
 |||||  
 Db 1347 AGAGGAGAGCTCATCTTCTTCAAGAAAGTGGAGCTGGAGACCTGCTGTGATGCGGCGCT 1406  
 |||||  
 QY 1086 GCTGGGCCACCCAGATGAGACCGCGTGAACGCGCATCCCTTCACTTCAAGAGAGTGAAGCT 1145  
 |||||  
 Db 1407 GCTGGGCCACCCAGATGAGACCGCGTGAACGCGCATCCCTTCACTTCAAGAGAGTGAAGCT 1466  
 |||||  
 QY 1146 CCTAAGCATTAAGTGAAGTGAAGTCTTACCCACAAAGTTACCCCGAGTGTGATCCAGCA 1205  
 |||||  
 Db 1467 CCTAAGCATTAAGTGAAGTGAAGTCTTACCCACAAAGTTACCCCGAGTGTGATCCAGCA 1526  
 |||||  
 QY 1206 CTTGGGCTACTCTTCTTCAAGATGAGCCCTGAGGACATTTGGCAAGTGAATGTGACCTC 1265  
 |||||  
 Db 1527 CTTGGGCTACTCTTCTTCAAGATGAGCCCTGAGGACATTTGGCAAGTGAATGTGACCTC 1586  
 |||||  
 QY 1266 CCTGAGAGACCTGAGAGGCTTGTGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1325  
 |||||  
 Db 1587 CCTGAGAGACCTGAGAGGCTTGTGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1641  
 |||||  
 QY 1326 TCTCTGGGGGGGCTCCACAGAGTGGGCGACCGCTGATGACCGCTTTGTGAAGGAGGAGG 1385  
 |||||  
 Db 1642 -----CAGGTGGCGCACCCCTGATGACCGCTTTGTGAAGGAGGAGGAGG 1682  
 |||||  
 QY 1386 CCAGCTAGACAAAGACACCTTGAAGACCCCTTGAACCGCTTGAACCTGGGTACCTGTGCTC 1445  
 |||||  
 Db 1683 CCAGCTAGACAAAGACACCTTGAAGACCCCTTGAACCGCTTGAACCTGGGTACCTGTGCTC 1742  
 |||||  
 QY 1446 CCTGAGGCGGAGGAGCTGAGCTCGTCCCGCCAGAGAGCATCTGGGGGGGTGAGGCGCCCA 1505  
 |||||  
 Db 1743 CCTGAGGCGGAGGAGCTGAGCTCGTCCCGCCAGAGAGCATCTGGGGGGGTGAGGCGCCCA 1802  
 |||||  
 QY 1506 GGACCTGAGACGTTGTGAACCCAGGACGCTGACGCTCTTATCCCAAGGCGCGGCTTGC 1565  
 |||||

Db 1803 GGACCTGAGACGTTGTGAACCCAGGACCTGAGAGCTCTTATATCCAAAGGCGCGGCTTGC 1862  
 |||||  
 QY 1566 TTTCAGACATGAACGGGTCCGAATACTTGTGAGATCCAGTCTTCTGTGGTGGGGC 1625  
 |||||  
 Db 1863 TTTCAGACATGAACGGGTCCGAATACTTGTGAGATCCAGTCTTCTGTGGTGGGGC 1922  
 |||||  
 QY 1626 CCCCAGGAGATTGAAGGGGCTCAGTACAGATGTGACATGTGACTTGGGACGTT 1685  
 |||||  
 Db 1923 CCCCAGGAGATTGAAGGGGCTCAGTACAGATGTGACATGTGACTTGGGACGTT 1982  
 |||||  
 QY 1686 CATGAGCTGGGAGCGGATGCGGTGCTCCCTTGAATGTGCTGAGTGAAGTGAAGAACTTCT 1745  
 |||||  
 Db 1983 CATGAGCTGGGAGCGGATGCGGTGCTCCCTTGAATGTGCTGAGTGAAGTGAAGAACTTCT 2042  
 |||||  
 QY 1746 GGGACCCACGTGAGAGGCTCTGAAGGCGGAGAGCGGACCGCGGCTGGGAGTGGAT 1805  
 |||||  
 Db 2043 GGGACCCACGTGAGAGGCTCTGAAGGCGGAGAGCGGACCGCGGCTGGGAGTGGAT 2102  
 |||||  
 QY 1806 CCTAGGCGAGCGGAGAGGACGACCTGAGACGCTGGGAGCTGAGGCGGACATCC 1865  
 |||||  
 Db 2103 CCTAGGCGAGCGGAGAGGACGACCTGAGACGCTGGGAGCTGAGGCGGACATCC 2162  
 |||||  
 QY 1866 CAAGGCTACCTGCTTCTAGACCTCAGCGTCAAGAGACCTCTCGGGAGCGGCTGCT 1935  
 |||||  
 Db 2163 CAAGGCTACCTGCTTCTAGACCTCAGCGTCAAGAGACCTCTCGGGAGCGGCTGCT 2222  
 |||||  
 QY 1926 CCTAGAGACTGAGACTTCTTCCACCGTCTTGGGACCTGCTTCTAGCTTCCACCTGCGCTG 1985  
 |||||  
 Db 2223 CCTAGAGACTGAGACTTCTTCCACCGTCTTGGGACCTGCTTCTAGCTTCCACCTGCGCTG 2282  
 |||||  
 QY 1986 AGGGCCCACTCCCTTCTGTCGCCCCAGCCCTGCTGGGGATCCCGGCTGGGAGGAGAG 2045  
 |||||  
 Db 2283 AGGGCCCACTCCCTTCTGTCGCCCCAGCCCTGCTGGGGATCCCGGCTGGGAGGAGAG 2342  
 |||||  
 QY 2046 GCAGGGTGAATCCCGTTCACCCCAAGAGAACTCGGCTAGTAAACGGGAACATGCC 2105  
 |||||  
 Db 2343 GCAGGGTGAATCCCGTTCACCCCAAGAGAACTCGGCTAGTAAACGGGAACATGCC 2402  
 |||||  
 QY 2106 CTTGAGACAAAAAATAAAAAAAAAAAAAA 2136  
 |||||  
 Db 2403 CTTGAGACAAAAAATAAAAAAAAAAAAAA 2433  
 |||||

RESULT 7  
 HUMMPF  
 LOCUS 2093 bp mRNA PRI 01-FEB-2000  
 DEFINITION Human mRNA for pre-pro-megakaryocyte potentiating factor, complete cds.  
 ACCESSION D49441  
 VERSION D49441.1 GI:1129078  
 KEYWORDS MRF; megakaryocyte potentiating factor.  
 SOURCE Homo sapiens pancreatic cancer cell line: HPC-Y5 cDNA to mRNA, clone: pK027.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 11 to 2093)  
 Kojima, T., Oh-beda, M., Hattori, K., Taniuchi, Y., Tamura, M., Ochi, N. and Yamaguchi, N.  
 Molecular cloning and expression of megakaryocyte potentiating factor cDNA  
 J. Biol. Chem. 270 (37), 21984-21990 (1995)  
 MEDLINE 95394969  
 REFERENCE 2 (bases 1 to 10)  
 Kojima, T.  
 JOURNAL Unpublished (1996)  
 REFERENCE 3 (bases 1 to 2093)  
 Kojima, T.  
 JOURNAL Direct Submission  
 TITLE Submitted (23-FEB-1995) to the DDBJ/EMBL/GenBank databases. Tetsuo Kojima, Chugai Pharmaceutical Co. Ltd., Fuji Golemba Research Laboratories; 1-135 Komakado, Golemba-shi, Shizuoka prefecture 412,







QY 520 CGCATCAGGAAGCCATATGTGACCTGCTCCCGAAGGGGGCTCCCGACGACAGCGCTG 579  
 |||||  
 Db 543 CGCATCAGGAAGCCATATGTGACCTGCTCCCGAAGGGGGCTCCCGACGACAGCGCTG 602  
 QY 580 CTGCTGCGGCTCTGCGCTGCTGGGGGTGTGGGGGGTCTCTGCTGAGGAGGCTGATGTG 639  
 |||||  
 Db 603 CTGCTGCGGCTCTGCGCTGCTGGGGGTGTGGGGGGTCTCTGCTGAGGAGGCTGATGTG 662  
 QY 640 CGGCTCTGGGAGGCTGCTGGCTTGGACCTGCTGGGCTTGTGGGGAGCTGGGCGAA 699  
 |||||  
 Db 663 CGGCTCTGGGAGGCTGCTGGCTTGGACCTGCTGGGCTTGTGGGGAGCTGGGCGAA 722  
 QY 700 GTGCTGTACCCCGGCTGTGAGCTGCCGGGAGCCCTGAGCAGAGGACAGAGAGCA 759  
 |||||  
 Db 723 GTGCTGTACCCCGGCTGTGAGCTGCCGGGAGCCCTGAGCAGAGGACAGAGAGCA 782  
 QY 760 GCCAGGGGGCTCTGCAAGGGGGGGGAGACCCCTACAGGCCCCCGTGCACATGCTGTG 819  
 |||||  
 Db 783 GCCAGGGGGCTCTGCAAGGGGGGGGAGACCCCTACAGGCCCCCGTGCACATGCTGTG 842  
 QY 820 TCCAGATGAGACGCTCTGCGGGGCTGCTGGCCGCTGGGCGCAAGCCATATCCGAGC 879  
 |||||  
 Db 843 TCCAGATGAGACGCTCTGCGGGGCTGCTGGCCGCTGGGCGCAAGCCATATCCGAGC 902  
 QY 880 ATCCCGCAGGGCATGTGCGCGCTGGCGGCAACGCTCTCTCGGGACCATCTGCGCG 939  
 |||||  
 Db 903 ATCCCGCAGGGCATGTGCGCGCTGGCGGCAACGCTCTCTCGGGACCATCTGCGCG 962  
 QY 940 CAGCCTGAAGGAGCATCTCTCGGCGCGGCTTCCGGGGGAGAGTGAGCAACAGCCCTG 999  
 |||||  
 Db 963 CAGCCTGAAGGAGCATCTCTCGGCGCGGCTTCCGGGGGAGAGTGAGCAACAGCCCTG 1022  
 QY 1000 CCTTCAGGCAAGAAAGCCCGGAGATAGACAGAGCCCTATCTTCTACAGAAAGTGGAG 1059  
 |||||  
 Db 1023 CCTTCAGGCAAGAAAGCTCCCGAGATAGACAGAGCCCTATCTTCTACAGAAAGTGGAG 1082  
 QY 1060 CTGGAAGCCTCGTGATGATGCGGCGCTGTGGCCACCCAGATGAGACCGGCTAAGCCATC 1119  
 |||||  
 Db 1083 CTGGAAGCCTCGTGATGATGCGGCGCTGTGGCCACCCAGATGAGACCGGCTAAGCCATC 1142  
 QY 1120 CCTTCACCTACGAGACGCTGAGCTGCTTAAGCATAACTGATGAGCTTACCCACAA 1179  
 |||||  
 Db 1143 CCTTCACCTACGAGACGCTGAGCTGCTTAAGCATAACTGATGAGCTTACCCACAA 1202  
 QY 1180 GGTATACCCGAGTGTGATCAGACACCTGGGCTACCTTCTCAAGATGAGCCCTGAG 1239  
 |||||  
 Db 1203 GGTATACCCGAGTGTGATCAGACACCTGGGCTACCTTCTCAAGATGAGCCCTGAG 1262  
 QY 1240 GACATTCGCAAGTGGATGTACGCTCCCTGAGAGACCTGAAAGGCTTGTGAAAGTGCAC 1299  
 |||||  
 Db 1263 GACATTCGCAAGTGGATGTACGCTCCCTGAGAGACCTGAAAGGCTTGTGAAAGTGCAC 1322  
 QY 1300 AAAGGCAAGAAATGAGTCTCAGAGCTCCTGGGCGGCCCTCCCAAGAGTGGCCACCTG 1359  
 |||||  
 Db 1323 AAAGGCAAGAAATGAGTCTC-----CAGGTGGCCACCTG 1388  
 QY 1360 ATGACCGCTTTGTGAAGGGAAGGGGCCAGCTAGACAAAGACACCTAGACACCTGACC 1419  
 |||||  
 Db 1359 ATGACCGCTTTGTGAAGGGAAGGGGCCAGCTAGACAAAGACACCTAGACACCTGACC 1418  
 QY 1420 GCCCTTACCTGGGTACCTGTCTCCCTACGCCCGAGAGCTGAGCTCGTGCCGCCCC 1479  
 |||||  
 Db 1419 GCCCTTACCTGGGTACCTGTCTCCCTACGCCCGAGAGCTGAGCTCGTGCCGCCCC 1478  
 QY 1480 AGCAGCATCTGGGGGAGGAGGCGCCAGACCTGAGACAGCTGAGCCCAAGGACCTGAGC 1539  
 |||||  
 Db 1479 AGCAGCATCTGGGGGAGGAGGCGCCAGACCTGAGACAGCTGAGCCCAAGGACCTGAGC 1538  
 QY 1540 GTCCCTATCCCAAGGCGCGCTGCTTTCAGAAACATGAACGGGCTCGATACTTGTG 1599  
 |||||  
 Db 1539 GTCCCTATCCCAAGGCGCGCTGCTTTCAGAAACATGAACGGGCTCGATACTTGTG 1598  
 QY 1600 AAGATCCAGTCTCTTCTGCTGGGTGGGCGCCCAAGAGGATTTGAAGGCGCTCAGTCAGCAG 1659

Db 1599 AAGATCCAGTCTCTCTGCTGGGTGGGCGCCCGACGAGAGATTGAAGGCGCTCAGTCAGCAG 1658  
 QY 1660 AATGTGACATGAGCACTTGGCCACGCTTCAATGAAGCTGGCGAGAGGATGGGCTCTCGCTG 1719  
 |||||  
 Db 1659 AATGTGACATGAGCACTTGGCCACGCTTCAATGAAGCTGGCGAGAGGATGGGCTCTCGCTG 1718  
 QY 1720 ACTGTGCTGAGGTGACAGAACTTCTGGGACCCACAGCTGAGAGGCTGAAAGCGGAGAG 1779  
 |||||  
 Db 1719 ACTGTGCTGAGGTGACAGAACTTCTGGGACCCACAGCTGAGAGGCTGAAAGCGGAGAG 1778  
 QY 1780 CGGCACCGCCCGGTGCGGAGTGTATCTACGCAAGCGGCGAGACAGACTGAGACGCTG 1839  
 |||||  
 Db 1779 CGGCACCGCCCGGTGCGGAGTGTATCTACGCAAGCGGCGAGACAGACTGAGACGCTG 1838  
 QY 1840 GGGCTGGGGCTACAGAGGGGGGATCCCGACAGCGGCTACCTGCTCTAGACTAGAGCTGCA 1899  
 |||||  
 Db 1839 GGGCTGGGGCTACAGAGGGGGATCCCGACAGCGGCTACCTGCTCTAGACTAGAGCTGCA 1898  
 QY 1900 GAGACCTCTGCGGAGAGCCCTGCTCTAGAGACTGAGACTGTTCACCGCTGCGCA 1959  
 |||||  
 Db 1899 GAGGCGCTCTGCGGAGAGCCCTGCTCTAGAGACTGAGACTGTTCACCGCTGCGCA 1958  
 QY 1960 CTGCTCTAGCTCCACCTGCTGCTGAGGCGCCCACTCTCTTGTGGCCCGAGCCCTGCT 2019  
 |||||  
 Db 1959 CTGCTCTAGCTCCACCTGCTGCTGAGGCGCCCACTCTCTTGTGGCCCGAGCCCTGCT 2018  
 QY 2020 GGGGATCCCGGCTGCGGAGAGGAGGAGGAGGATGATCCCGCTTCCAGCCCGCAAGAGACT 2079  
 |||||  
 Db 2019 GGGGATCCCGGCTGCGGAGAGGAGGAGGAGGATGATCCCGCTTCCAGCCCGCAAGAGACT 2078  
 QY 2080 CGCGCTCAGTAAAGGGAACATGCCCTCGCAGACAAAAA 2138  
 |||||  
 Db 2079 CGCGCTCAGTAAAGGGAACATGCCCTCGCAGCAAAAAA 2137  
 RESULT 9  
 AF180951  
 LOCUS 1195 bp mRNA 16-JAN-2001  
 DEFINITION Homo sapiens mesothelin/megakaryocyte potentiating factor mRNA,  
 partial cds.  
 ACCESSION AF180951  
 VERSION AF180951.1 GI:6014586.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1195)  
 AUTHORS Scholler,N., Fu,N., Yang,Y., Ye,Z., Goodman,G.E., Hellstrom,K.E.  
 and Hellstrom,I.  
 TITLE Soluble member(s) of the mesothelin/megakaryocyte potentiating  
 factor family are detectable in sera from patients with ovarian  
 carcinoma  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11531-11536 (1999)  
 MEDLINE 99432267  
 PUBMED 10500211  
 REFERENCE 2 (bases 1 to 1195)  
 AUTHORS Scholler,N., Yang,Y. and Hellstrom,I.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-AUG-1999) Tumor Immunology, PNRI, 720 Broadway Rm.  
 511, Seattle, WA 98107, USA  
 FEATURES  
 source  
 1..1195  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 <1..>1195  
 /codon\_start=1  
 /product="mesothelin/megakaryocyte potentiating factor"  
 /protein\_id="AA01409.1"  
 /db\_xref="GI:6014587"  
 /translation="VERTACPSGSKKAREIDESLIFYKKWLEACVDALLATOMDRVN  
 AIPFYEQDLVLEKHLDELTPQGPESVIOHLGYLFLKMSPEDIRKMNVTSLTAL



```

LEVNKHEMSPOVALTIDRFVKRGOLDKIDTLDTLTAFAYPGLCSPEELSSVPPSS
IMAVRPODDETCBPRLDVLVYKPARILAFLOMNSSEVYKIOSFAGAPTEDLALSOO
NMSMDLATEFKRLRDAVLPLTVAEVOKLFPHEVJKAERHPRVMIIRODDID
TILGLOGGIPNGLYVLDSVOGRGDAAGAGRAGVEGALSHPSLCGRPIGDALP
PRTWCSHRBGTAPSLHPGLRALPCWPOKWSPPQEOBARVIVPPEOENSHSVNEN
MPPADT"
BASE COUNT      235 a      382 c      373 g      205 t
ORIGIN
Query Match      45.98; Score 980.8; DB 9; Length 1195;
Best Local Similarity 91.18; Pred. No. 3.8e-164;
Matches 1108; Conservative 0; Mismatches 2; Indels 106; Gaps 2;

OY  982 GTGAGAGAGACAGGCTGCTTCACGGCAAGAGCCCGCGAGATAGAGAGGCTCATC 1041
    1 GTGAGAGAGACAGGCTGCTTCACGGCAAGAGCCCGCGAGATAGAGAGGCTCATC 60
OY  1042 TTCTACAGAGAGTGGAGCTGGAAGCCCTGCGTGGATGGCGCCCTGCGCCACAGATG 1101
    61 TTCTACAGAGAGTGGAGCTGGAAGCCCTGCGTGGATGGCGCCCTGCGCCACAGATG 120
OY  1102 GACCGCGTGAAGCCGATCCCTTACCTACGAGACGCTGACGTCCTTAAGCATAACTG 1161
    121 GACCGCGTGAAGCCGATCCCTTACCTACGAGACGCTGACGTCCTTAAGCATAACTG 180
OY  1162 GATGAGCTCTACCCACAGGTTACCCGAGTCTGTGATCCAGACGCTGGCTACTCTTC 1221
    181 GATGAGCTCTACCCACAGGTTACCCGAGTCTGTGATCCAGACGCTGGCTACTCTTC 240
OY  1222 CTCAAGATGAGCCCTGAGGACATTCGCAAGTGAATGAGCTGCTCCGAGACCCCTGAG 1281
    241 CTCAAGATGAGCCCTGAGGACATTCGCAAGTGAATGAGCTGCTCCGAGACCCCTGAG 300
OY  1282 GCTTTCGTTGAAGTCGACAAAGGGCAGAGAAATGAGTCTCAGGCTCCTCGCGGCCCTC 1341
    301 GCTTTCGTTGAAGTCGACAAAGGGCAGAGAAATGAGTCTC----- 339
OY  1342 CCACAGGTGGCCACCTGATGACCGCTTTGTGAAGGAGAGGGCCAGCTAGACAAAGAC 1401
    340 ---CAGGTGGCCACCTGATGACCGCTTTGTGAAGGAGAGGGCCAGCTAGACAAAGAC 396
OY  1402 ACCCTAGACACCTGACCGGCTTCTACCCCTGGTACCTGTGCTCCCTCAGCCCGAGAG 1461
    397 ACCCTAGACACCTGACCGGCTTCTACCCCTGGTACCTGTGCTCCCTCAGCCCGAGAG 456
OY  1462 CTGAGCTCCGCGCCCGCCAGAGCATCTGGGCGGTCAAGGCCCGAGACCTGACACGCT 1521
    457 CTGAGCTCCGCGCGCCCGCCAGAGCATCTGGGCGGTCAAGGCCCGAGACCTGACACGCT 516
OY  1522 GACCAAGGACAGCTGAGAGCTCTATCCCAAGGCCCGCTTGGTTCAGACATGAAAC 1581
    517 GACCAAGGACAGCTGAGAGCTCTATCCCAAGGCCCGCTTGGTTCAGACATGAAAC 576
OY  1582 GGGTCCGAATACCTTGTGAAGATCCAGTCTTCTGGGTGGGGCCCGACGAGAGATTG 1641
    577 GGGTCCGAATACCTTGTGAAGATCCAGTCTTCTGGGTGGGGCCCGACGAGAGATTG 636
OY  1642 AAGGCGCTCATATCAGCAGAAATGTAGATGACTTGGCCACGTTCTATGAACCTGCGAGC 1701
    637 AAGGCGCTCATATCAGCAGAAATGTAGATGACTTGGCCACGTTCTATGAACCTGCGAGC 696
OY  1702 GATCGGCTGCTGCGGCTGACTGCTGAGTGCAGAAACTTCTGGGAGCCCGACAGTGGAG 1761
    697 GATCGGCTGCTGCGGCTGACTGCTGAGTGCAGAAACTTCTGGGAGCCCGACAGTGGAG 756
OY  1762 GGCGTGAAGCGGAGAGAGCGGCGCCGCGGTGCGGAGCTGGATCTTACGCGAGCGGAG 1821
    757 GGCGTGAAGCGGAGAGAGCGGCGCCGCGGTGCGGAGCTGGATCTTACGCGAGCGGAG 816
OY  1822 GAGACGCTGAGACAGCTGGGGCTGAGGGCTACAGGGCGGCGATCCCAACGGCTACTGCTC 1881
    817 GAGACGCTGAGACAGCTGGGGCTGAGGGCTACAGGGCGGCGATCCCAACGGCTACTGCTC 876

```

```

OY  1882 CTAGACCTCAGCGTGCA----- 1898
    1882 CTAGACCTCAGCGTGCA----- 1898
OY  877 CTAGACCTCAGCGTGCAAGAGTGGCGGGCGGCCAGGCCAGGCGTGGGGCAGAGCTGG 936
    877 CTAGACCTCAGCGTGCAAGAGTGGCGGGCGGCCAGGCCAGGCGTGGGGCAGAGCTGG 936
OY  1899 -----AGAGACCTCTCGGGGAGCGCC 1919
    1899 -----AGAGACCTCTCGGGGAGCGCC 1919
OY  937 GCGGTGAGAGTGGGGGCTCTGAGTACACCCCTCTCTCTGTGAGAGGGCCCTCTCGGGAGCGCC 996
    937 GCGGTGAGAGTGGGGGCTCTGAGTACACCCCTCTCTCTGTGAGAGGGCCCTCTCGGGAGCGCC 996
OY  1920 CTGCGCTCTAGAGACCTGGACCTGTTCTACCGTCTCTGACCTGCTCTTCTCTGACCTT 1979
    1920 CTGCGCTCTAGAGACCTGGACCTGTTCTACCGTCTCTGACCTGCTCTTCTCTGACCTT 1979
OY  997 CTGCGCTCTAGAGACCTGGACCTGTTCTACCGTCTCTGACCTGCTCTTCTCTGACCTT 1056
    997 CTGCGCTCTAGAGACCTGGACCTGTTCTACCGTCTCTGACCTGCTCTTCTCTGACCTT 1056
OY  1980 GCGCTGAGAGGCGCCACGCTGCTGCGGCGCCAGACGCTGCGGGATCCCGCTGGAGCCAG 2039
    1980 GCGCTGAGAGGCGCCACGCTGCTGCGGCGCCAGACGCTGCGGGATCCCGCTGGAGCCAG 2039
OY  1057 GCGCTGAGAGGCGCCACGCTGCTGCGGCGCCAGACGCTGCGGGATCCCGCTGGAGCCAG 1116
    1057 GCGCTGAGAGGCGCCACGCTGCTGCGGCGCCAGACGCTGCGGGATCCCGCTGGAGCCAG 1116
OY  2040 GAGCAGGCGAGGCTGATCCCGCTTCCACCCCAAGAGAACTGCGGCTCAGTAAAGGGAGAC 2099
    2040 GAGCAGGCGAGGCTGATCCCGCTTCCACCCCAAGAGAACTGCGGCTCAGTAAAGGGAGAC 2099
OY  1117 GAGCAGGCGAGGCTGATCCCGCTTCCACCCCAAGAGAACTGCGGCTCAGTAAAGGGAGAC 1176
    1117 GAGCAGGCGAGGCTGATCCCGCTTCCACCCCAAGAGAACTGCGGCTCAGTAAAGGGAGAC 1176
OY  2100 ATGCCCCCTGCAGACA 2115
    2100 ATGCCCCCTGCAGACA 2115
OY  1177 ATGCCCCCTGCAGACA 1192
    1177 ATGCCCCCTGCAGACA 1192

RESULT 10
D86370
LOCUS      D86370      2135 bp      mRNA      ROD      07-FEB-1999
DEFINITION mouse mRNA for megakaryocyte potentiating factor, complete cds.
ACCESSION D86370.1 GI:1408516
VERSION    D86370.1
KEYWORDS   megakaryocyte potentiating factor.
SOURCE     Mus musculus tissue_11b:Lung cDNA to mRNA.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2135)
Kojima,T., Taniguchi,Y., Hattori,K. and Oh-eda,M.
Submitted (03-JUL-1996) to the DDBJ/EMBL/Genbank databases. Tetsuo
Kojima, Chugai Research Institute for Molecular Medicine, Inc.,
Cytochrome Research Program, 153-2, Nagai, Mihari, Ibaraki 300-41,
Japan (E-mail:kojimatemb.infoweb.or.jp, Tel:0298-30-6211,
Fax:0298-30-6270)
2 (bases 1 to 2135)
Kojima,T., Taniguchi,Y., Hattori,K. and Oh-eda,M.
mouse Megakaryocyte Potentiating Factor cDNA
Unpublished (1996)
REFERENCE
3 (sites)
Kojima,T., Oh-eda,M., Hattori,K., Taniguchi,Y., Tamura,M., Ochi,N.
and Yamaguchi,N.
Molecular cloning and expression of megakaryocyte potentiating
factor cDNA
JOURNAL    J. Biol. Chem. 270 (37), 21984-21990 (1995)
MEDLINE    95394969
FEATURES
source
location/Qualifiers
1..2135
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_lib="Lung"
109..255
/codon_start=1
/product="megakaryocyte potentiating factor"
/protein_id="BAI3077.1"
/db_xref="GI:1408517"
/translation="MALPAPRLLPGSGSPICRSRLLLLSGLMIPRLTOTRTKTSQ
EATLLHAVNGADPASPGLTGLTVEEYSDSMEGAKLAAVAVRKNITLGGHQR
CLARRLPRLHLDDELNALPDLILLFLNPAWPGQAQAHFSPISIKRANVDLPRLSL

```

ROBLMEALCOCGVYGFVSEADYRALGLACDLPGKVARSSVLLPMLAGCGPLD  
QSOEKAVERLRSRTQYGPSSKMSVSTLALOSLVLDLIVQSIYQSPKVKAEMLQH  
ISRPRLSGKLYIHPFRPDAEOKAKPEYKVEDLIEFOMNEPACVGTML  
AROMDIYNEIPEFYEOISIFKHKLIDKYPOGYSPLIOOQGHPRFYSPLDIHOMNT  
SPDVTYLLKLVSKQKNNADALIALVCTLYNGGSLDEDMKALADILPLSLCDSPOD  
LHVSPPVMMVLVGPQDLKCSORHLGLVYKACSAFONVGLFEKIKTFLLGASVK  
DLRALSOHNVSMIDATFKRLQVDSLVGLSAEYOKLGPINVIDKTEDSVPYDMIF  
ROHKDIDRLGLGIGGIPINGYIVLDENVEAREAFSSRASLLGPGFVLIWIPALLPALRL  
S<sup>+</sup>

polya\_signal 2089. 2094  
BASE COUNT 504 a 591 c 586 g 454 t  
ORIGIN

Query Match 42.6%; Score 910; DB 10; Length 2135;  
Best Local Similarity 67.8%; Pred. No. 1.1e-151;  
Matches 1415; Conservative 0; Mismatches 620; Indels 53; Gaps 8;

QY 69 GGTGCTCCCTCCCTGGGATCTACACAGACATGGCCCTTG-CACGGCTCGACCCCTGTT 127  
DB 78 GGGACCTCAGAGTCAATGTTATCCACAGACATGGCTTGCAACAGCTCGACCCCTGCT 137  
QY 128 GGTCTGTGGGGA-----CCGCCCTGGCAGCTCTGTTCCCTGCTCTTACGCTCGG 179  
DB 138 GGGGTCTGTGTGAAGTCCCATCTGCAGCCGAAGCTTCTACTGCTTCTTAACTTGG 197  
QY 180 ATGGGTGATCCCGGAGAGCCCTGGCTGAGAGACAGAGAGAGTGTGCCCCCTGGG 239  
DB 198 GTGATACACAGCTGTGCAGACCCAGACTACAAAGAACAGAGAGGCGACACTCTCCA 257  
QY 240 GGGATCTCTGACACCCCATTAACATTTCAGCTCTCCCTCGCACTCTTGGCTT 299  
DB 258 TGTCTG---TGAACGGTGGCGCTGACTTGGCAGTCTCCACAGAGCTCTTCTTGGCTT 314  
QY 300 CCCGTGTGGAGGTGTGGCGCTGAGACAGAGAGTGTCCGGAGGCTGGCTGTGGCTT 359  
DB 315 CACATGTAGAGAGTATGTAGCTGAGCATGGAACCAAGGGGCTGGCTATGGCTGT 374  
QY 360 GGCACAGAAATGTCAAGCTCTCAACAGACAGCTGTGCTGTGCTCAACGGCTCTC 419  
DB 375 AAGACAGAAAGAACTTACACTCCGGGAGACATCAGCTGGTGTCTGGACAGCTGGCTTCC 434  
QY 420 TGAGC-----CCCCGAGACCTGAGACGCTTCCCATTTGAAGCTGTGTAATTCCTCAA 473  
DB 435 TAGGACCTCACCGAGAGAACTGAATGCTTCTCAGCTGAGACGCTGCTTCTCTCAA 494  
QY 474 CCCAGATCCGTTTCGGGGCCCCCAGGCTGCACCCGTTTCTTCTCCCGCATACAGCAAGC 533  
DB 495 CCCAGCATGTTTCCAGGGCAACAGGCTGTGCCCCCATCTTCTCCCATCTCTTAAAGC 554  
QY 534 CAATGTGACCTGTCCCGAGGGGGCTCCCGAGAGACAGGAGGCTGTGCTCGGGCTCT 593  
DB 555 CAATGTGATGTACTACCCAGAGAGTCTCTGAGAGGCGACAGAGCTGTGATGAGAGCTCT 614  
QY 594 GGCCTGCTGGGGGTGGGGGGTCTCTCTGAGCGAGCTGATGTGGGGCTCTGGAGG 653  
DB 615 GAAGTGGCCAGGGCGTGTGATTTCAAGTAGTAGAGAGCAATGTGGCGGCTCTGGAGG 674  
QY 654 CCTGGCTTGGACCTGCTTGGGGCGCTTGTGGCGAGTGGCGGCAAGTGTGCTTACCCG 713  
DB 675 CCTGGCTTGTACCTGCTGGGAAATTTGTGGCCAGATCTTCCGAGATTTCTCTCCCTG 734  
QY 714 GCTGTGAGTGTCCCGGAGACCTTGGACAGACAGACAGAGAGGAGGAGGCGGCTCT 773  
DB 735 GCTGGCAGAGATGCAACACCCCTGGACAGAGCCAGGAAAGAGAGTCAAGGAGGTTCT 794  
QY 774 GCAGGGCGGGGAGACCCCTTCAGGGCCCCCGCTGCATGTGTCTCTCCAGATGAGAGC 833  
DB 795 GAGGAGTGAAGAACCAATATGCGCCCAATCGAAGTGTGCTACACCTGTGATGC 854  
QY 834 TCTGTGGGGGCTGTGCGCCGCTGTGGGCAAGCCATCATCCGACAGATCCCGCAGGAGT 893  
DB 855 CCTGAGAGCTTGTGAGCATGTTGATGATGATCCATGCTCCAGAGACATCCCAAGAGTGT 914

QY 894 CTTGGCCCGCTGGGGCAACCGCTCTCTCAGGACCATCTGGGGGAGCTGAGCGAGC 953  
DB 915 CAAGCTCAATATGGCTGCAACATCTCCAGAGACCCCTCCAGGCTGGGGTCTTAAGCTGAC 974  
QY 954 CATCTCTCGGGCGGGTCTCCGGGGGAGAGTGAAGAGACAGCTGTCTCTCAGCAGAA 1013  
DB 975 GGTATACACCAAGGTTCCAGGGAGATGACAGAACAGAACTGTCCCTCCAGGAAAGGA 1034  
QY 1014 GGGCCGGAGATTAAGAGAGAGCTCATCTCTACAAAGAGTGGGGAGCTGGAAGCTGGCT 1073  
DB 1035 GCGCTACAGAGTGAATGAAGACATCTCTACAGATTTGGAGAGCTGTGTGT 1094  
QY 1074 GGATGCGGCGCTGTGGCGCACCCAGATGACCGGCTGAAGCCCATCTCCCTACCTACGA 1133  
DB 1095 AGATGGACACATGCTGGCGCAGACAAATGAGACTGTGAGACAGATTTCCCTTACCTTGA 1154  
QY 1134 GCAGCTGACCTCTTAAGCATTAACCTGATGAGCTTACCCACAAAGTTTACCCGAGTC 1193  
DB 1155 GCAGCTCAGTATCTTTAAGCAACAACTGGACACAGCACTACCCACAGGCTATCTGAGTC 1214  
QY 1194 TGTGATCCAGACCTGGGCTACCTCTCTCTCAAGATGAGCCCTGAGGACATTTGCAAGTG 1253  
DB 1215 CCGTATCCAGAGCTGTGGCTCACTTCTTCAATGTGTAGCCCTGAGACATTCACAGAGTG 1274  
QY 1254 GAATGTGACGTCCCTGAGACCTGTAAGCTTGTGCTGAGTGCACAAAGGGCAGAAAT 1313  
DB 1275 GAATGTGACGTCCACAGACAGTGAATAAGTCTGTCTAAAGTCAGCAAAAGGACAAAGAT 1334  
QY 1314 GAGTCTCAGGCTCTCGGGGGCCCTCCACAGGTGGCCACACCTGATGACAGCTTTGT 1373  
DB 1335 GAATGCTCAGGC-----GATTCCTGTGGTGGCTGTATCT 1370  
QY 1374 GAAGGGAAGGGGCGAGTACAAAGACACCTGACACCCCTGACCGCTTCTACCTGG 1433  
DB 1371 TCGGGAGAGAGGCGAGCTGTGAGAGAGATGTGTAAAGCCCTGGGCGATTCCTTAAAG 1430  
QY 1434 GTACTGTCTCCCTGACACCCCGAGAGCTGAGTCCCTGCCCCCAGACAGATCTGGGC 1493  
DB 1431 CTATCTATGTGACTTACACCCCGAGAGATCTGACACTGGTACCTCTCACTGATGTGCT 1490  
QY 1494 GGTACGGCCCCAGACCTGTGACACGTGTGACCCAAAGCAGCTGTCTTATCCAA 1553  
DB 1491 GGTGTGGGCCCCAAGACCTGTGACAACTGACGAGAGGAGATGTGGGTCTCTCTACCAAA 1550  
QY 1554 GGGCCGCGCTGTTCGAACATGAACGAGTCCGAATCTTGTGAAGATCCAGTCTT 1613  
DB 1551 GGCCTGCTCAGCTTCCGAATGTGAGCGGCCCTGAGAACTTTGAGAAATCAAGACATT 1610  
QY 1614 CTTGGGTGGGGCCCCCAGAGAGATTTGAAGCGCTCTGAGTACAGATGTGAGCATGGA 1673  
DB 1611 CTTGGGTGGGGCTTCGTAAGAGACCTCGGGCCCTCAGCCAGCAGATGTGAGCATGGA 1670  
QY 1674 CTTGGCCACGTTACGAAGCTGTGAGAGCGATGTGCTGTGCTGTGACTGTGTGAGGT 1733  
DB 1671 CATACCACTTTCGAAGAGCTGAGAGTGTGCTGTGAGGTGTGAGGTGTGAGGT 1730  
QY 1734 GCAGAAATCTTGGGAGACCCAGTGAAGGCTGGAAGCGGAGAGAGCGGACCGGCT 1793  
DB 1731 ACAGAAATCTTGGGGGCCCAACATTTGTGAGCTGGAAGCAGAGAGATTAAGACCTGT 1790  
QY 1794 GCGGAGTGTATCTTACGGCAGAGCGGAGACGACTGAGACAGCTGTGGGGCTTACGA 1853  
DB 1791 CCGTACTGTGGCTTCGGGAGCATCAAGAAAGCTTACAGAGGCTGGGTTGGGACTTCA 1850  
QY 1854 GGGGGGATCCCAAGGCTACTGTGTCTTACCTTACGCTGTGCAAGAGACCTCTCGGG 1913  
DB 1851 GGGTGTATTCCTCAATGTGCTGTGCTGTGCTTCAATGTGCGAGAGGCTTCTCAG 1910  
QY 1914 GAGCGCTGCTCTTGAAGACCTGAGACCTGTCTCA-----CGGCTGAGAGCTGTCTGAGC 1970  
DB 1911 CAGAGCTTACTCTTGGGCCAGGATTTGTATTAATGATTTCCAGCTTGTCTTCCAGC 1970

QY	1971	CTCACCCTGGCGTAGAGGGCCCCCACTCCTCTCTCTCTGGCCCCAGCGCTGGGGAT	2025
Db	1971	TTTAAAGCTGAGCTAGACACACACCTCGCAAGGCTCTGTGGTCCAGCTCTAGTGGG66--	2028
QY	2026	CCCCGCTGGCGCAGAGGAGGACAGGGGTGATCCCGTTCACCCCAAGAGACTCGCGCT	2085
Db	2029	-CCCTCTTGACACAGAGAGTGGGACACAGGGGTCATTTGCCAAMAGTTTGAGAGACTCTTGAACT	2087
QY	2086	CAGTAAACGGGACATCGTCCCCCTGCACACAAAAA	2133
Db	2088	CAGTAAACAGTGGCATATGCTCCCTTGAAAAA	2135
RESULT	11		
LOCUS	D87351		
DEFINITION	Rattus norvegicus mRNA for mesothelin, complete cds.		
ACCESSION	D87351		
VERSION	D87351.1		
KEYWORDS	GI:10129877		
SOURCE	Erc: mesothelin. Rattus norvegicus (strain:Eker rat) renal carcinoma cell_line:ERC33 cDNA to mRNA.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (sites) Yamashta,Y., Yokoyama,M., Kobayashi,E., Takai,S. and Hino,O.		
TITLE	Mapping and determination of the cDNA sequence of the Erc gene preferentially expressed in renal cell carcinoma in the Tsc2 gene mutant (Eker) rat model		
JOURNAL	Biochem. Biophys. Res. Commun. 275 (1), 134-140 (2000)		
MEDLINE	20403886		
REFERENCE	2 (bases 1 to 2099)		
AUTHORS	Hino,O.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-AUG-1996) to the DDBJ/EMBL/GenBank databases. Okio Hino, Cancer Institute, Department of Experimental Pathology; Kam1-Ikebukuro 1-37-1, Toshima-ku, Tokyo 170-8455, Japan (E-mail:ohino@ms.u-tokyo.ac.jp, Tel:03-5394-3815, Fax:03-5394-3815)		
FEATURES			
SOURCE	location/Qualifiers 1..2099 /organism="Rattus norvegicus" /strain="Eker rat" /db_xref="taxon:10116" /cell_line="ERC33" /cell_type="renal carcinoma" 1..90 91..1968 /gene="Erc" 91..1968 /gene="Erc" /note="Homologue of megakaryocyte potentiating factor (MP)" /codon_start=1 /protein_id="BAB13512.1" /translation="MALPTAOPPLGSCGSPICSNPFLLLSLIGMLPDLQOTRTISO EALLIAVATGVDFVASTPTGFLGTLTDEVSGLSHGAKELAMAVROKNITLYQHQLR CLARLPKHLINELDALPLDLFLNPAMPGGQACANHPSLISKANVLPKRSLE RORLTALGKCGYGVGFVSEFTDAPGLAGLADLPGEFVAKSSVLLPYLARCGGPLD OGAKAVREVIKRSRGAPEPSTVSTVSLTALGLIYVDSYVHSYPKVITVLOG ISRPKSGKMTYTHPRFRDTRQKACPEKRENVNDENLIFYNLELEACVDGTLIL AGQMDLVNELIPTYEQLSIFKHRLDKTYPOQYPSLLIKQGHFRVYSPEDIRQWNVDT SPDIPVNLTKVSKGQKMDAOLYIACVYLRGSGALDDIDYKALDNIPLSYICDESPDD LHAIPSSYMLVGLHDLKDCSORIALVIGLYOKACGAFQNVNSGLEFEKIRIFELGASRE DLRALSOHNVSMOIAFERKLOVDALGVISAEVQKLGPHIGDKTEEDSKPVDWLF RQQCKDDSDGLGIGGGIPNYILLDFNVRAPESSGAPLIGPQVFAMIRALLSLALRL 5"		
3'UTR	1969..2099		
polyA_signal	2074..2079		

BASE COUNT	465 a	588 c	585 g	461 t	
ORIGIN					
Query Match	42.2%	Score 902.8:	DB 10:	Length 2099:	
Best Local Similarity	67.9%:	Pred. No. 2e-150:			
Matches 1400:	Conservative 0:	Mismatches 612:	Indels 50:	Gaps	
OY	69	GGTGCCTCCCTCCCTGGGATCTACACAGACATGGCTTGG-CAACGGCTCGACCCCTT	127		
Db	60	GGGACCTCAGAACCATTTGATTTCCACAGACATGGCTTGGCCACAGCCCAACCCCTGCT	119		
OY	128	GGTCTGTGGGGG-----CCGCCCTGGACCCCTGTTCTGCTTCGCTTCAGCCCTGG	179		
Db	120	GGGGTCTGTGGGAAGGCCCATCTGCAAGCCGACGACTTTCCTACTGCTTCTCTTAAGCTTGG	179		
OY	180	ATGGGTGCATCCGCGAGGACCTGTGCTGGAGACAGAGGAGCGAGTGTGCCCCCTGGG	239		
Db	180	GTGGTTGCCACTTCTGCGACCCGACGACTACAGAGCAACCCAGAGGCGCGCACTTCTCA	239		
OY	240	GGGAGTCTCTAGCAACCCCCCATTAACATTTCCAGCCCTGCTCCCTGCGCAACTCTTGGCT	299		
Db	240	TGCTG---TGACGGGACCCCTTACTTTCGAGCTTCTCCACAGGCTTCTTCTTGGCT	296		
OY	300	CCCGTGTGCGGAGGTGTCTGCGGCTTGAGACAGAGCTGTCCGGGAGTGGCTGTGGCTT	359		
Db	297	CACGTGTGATGAGGTATCTGGCTTAAGCATGGGACAGCGCAAGAGAGCTGGATATGCTGT	356		
OY	360	GGCACAAGAAATGTCAAGCTCTCACAGAGCAGCTGGCCTGTCTGGCTCACCGGCTCTC	419		
Db	357	GAGACGAGAAGATATCTGTCTCCAACTACTACGCTGCGCTGTGGGCCGCTGCGCTCC	416		
OY	420	TGAC-----CCCCGAGAGACTTGAGACGCTCCCATTTGAGACTGTCTATTCCTCA	473		
Db	417	TAAACACCTCACAAGAGGAGAACTGGATGTCTCCACATGGACCTGTGCTCTTCTCA	476		
OY	474	CCCAAGATGCTTCTGGGGGCCCAAGGCTGTACCCCTTCTCTCCGATACAGAGGC	533		
Db	477	TCCAGCCATGTTTCCGGGGGACAGGCTTGTGCCCACTTCTCTCCATCTCTAAAGC	536		
OY	534	CAATGTGAGACTGTCTCCGAGGGGGGCTCCGAGCACAGCGGCTGTGCTCGCGCTCT	593		
Db	537	CAATGTAAATGTACTCCACAGAGATCTCTGAGAGCCGCAAGAGCTGTACCGGGGCTCT	596		
OY	594	GCGCTGTGGGTGTGCGGGGGTCTCTGCTGAGCGAGCTGATGTGCGGGCTCTGGAGG	653		
Db	597	GAATGTCCAGAGGTGTGATGATTTCAAGTGATGTAGAGGATGACAGGCGCTCTCGAGG	656		
OY	654	CGTGGCTTGGGACCTGTCTGGGGGCTTGTGGCGGAGTGGGCGAAGTCTCTACCCGG	713		
Db	657	CTGTGCTGTACTGTCTGTGGGAATTCGTGGCCAAATTTGGGAAGTCTCTCTCCCTG	716		
OY	714	GCTGTGTAGCTGCCCCGGGACCCCTTGAGCCAGACAGAGGAGCGAGCGGGCTCT	773		
Db	717	GCTGGCAAGATGGGAGAGACCCCTTGAGCCAGGCGCAAGGCTGTACGAGGAGTTCT	776		
OY	774	GCAAGGGGGGGAACCCCTTACGCGCCCTGTGAATATGTTCTGTCTCCAGATGGAGGC	833		
Db	777	GAGAGGTGAGAGACCCCTTATGATGCTCCCTATGACTGATGTGTCTCCACCTCTGAATGC	836		
OY	834	TCTCTCGGGGCTGTCTGCTCCCTGTGGGCAAGCCATCATCTCCAGAGTCCGACAGGAT	893		
Db	837	CTGTGAGGGTTTGTGTATGTTGGATGAGTGCATTTGTCACAGATGCTCTTAAGATGT	896		
OY	894	CGTGGCGCGTGGCGGCAAGCTTCTCTCGAGCCCATCTTGGCGGCACTGGAAGGAC	953		
Db	897	TATCACTGAATGTGTGCAAGGCACTTCCAAAGAGCCCTCAGAGCTGGGTCTTAATGTGAC	956		
OY	954	CATCTCTCGGGCGGGGTCTCCGGGGGAGAGGAGCAAGCTGTCTTCAAGGAMA	1013		
Db	957	TGTACACACCCCAAGGTTCCGGCGGCAKAGACAGACAAAGGCTGCTCTCCAGGGAAGA	1016		
OY	1014	GGCCCCGAGATAGACGAGAGCTCATTTCTTCAAGAAAGTGGAGAGCTGAGGCTTGGCT	1073		







```
* 126888 126987: gap of 100 bp
* 126988 154616: contig of 27629 bp in length.
FEATURES
    source
        1..154616
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_1bp="RP11-728H8"
            /clone_11bp="RPC1-11 Human Male BAC"
            1..1098
                /note="assembly_fragment"
            1199..2218
                /note="assembly_fragment"
            2319..3383
                /note="assembly_fragment"
            3384..5809
                /note="assembly_fragment"
            5910..7395
                /note="assembly_fragment"
            7496..9344
                /note="assembly_fragment"
            9445..11953
                /note="assembly_fragment"
            12054..14571
                /note="assembly_fragment"
            14672..17477
                /note="assembly_fragment"
            17578..21206
                /note="assembly_fragment"
            21307..24213
                /note="assembly_fragment"
            clone_end:SP6
            vector_side:left"
            24314..29309
                /note="assembly_fragment"
            29410..33872
                /note="assembly_fragment"
            33973..40495
                /note="assembly_fragment"
            40596..48225
                /note="assembly_fragment"
            48326..54012
                /note="assembly_fragment"
            54113..61592
                /note="assembly_fragment"
            61693..71679
                /note="assembly_fragment"
            71780..77756
                /note="assembly_fragment"
            77857..85538
                /note="assembly_fragment"
            85639..96350
                /note="assembly_fragment"
            96451..110273
                /note="assembly_fragment"
            110374..126887
                /note="assembly_fragment"
            126988..154616
                /note="assembly_fragment"
            clone_end:T7
            vector_side:left"
BASE COUNT  31504 a 44661 c 46120 g 30028 t 2303 others
ORIGIN
Query Match      14.4%; Score 307.6; DB 2; Length 154616;
Best Local Similarity 82.4%; Pred. No. 1.2e-45;
Matches 402; Conservative 0; Mismatches 4; Indels 82; Gaps 1;
```

```
|||||
Db 15108 GGCAGAGAGCGGACCGCCGCGGAGACTGATCTTACGGCAGCGGAGGAGACT 15167
QY 1830 GGACACCTGGGGCTGGGGCTACAGGGGGGATCCCCCAAGGCTACCGTCTAGAGCT 1889
Db 15168 GGACACCTGGGGCTGGGGCTACAGGGGGGATCCCCCAAGGCTACCGTCTAGAGCT 15227
QY 1890 CAGCGTGCA----- 1898
Db 15228 CAGCATGCAAGTGGGGGGGGGGGGGCGGAGCCAGCGGCTGGGGGCGAGAGCTGGGGGCGTGA 15287
QY 1899 -----AGAGCCCTTGGGGAGCGCTGCTCC 1927
Db 15288 GGTGGGCGCTGTGAGTACACCCCTCTCTCTGTAGAGGCGCTCTCGGGGAGCGCTGCGCTCC 15347
QY 1928 TAGGACCTGGACCTGTCTCTACCGTCTGGACGACGCTCTAGAGCTCCAGCTGGGCTGAG 1987
Db 15348 TAGGACCTGGACCTGTCTCTACCGTCTGGACGACGCTCTAGAGCTCCAGCTGGGCTGAG 15407
QY 1988 GGGCCCACTCCCTTGTGGGGGGGGGGGGGGGATCGGCTGGGCTGGGCGGAGAGGAGCG 2047
Db 15408 GGGCCCACTCCCTTGTGGGGGGGGGGGGGGGATCGGCTGGGCTGGGCGGAGAGGAGCG 15467
QY 2048 ACGGGTATCCCGTTCACCCCAAGAGAGACTCCGCTCAGTAAACGGGAACATGCCCC 2107
Db 15468 ACGGGTATCCCGTTCACCCCAAGAGAGACTCCGCTCAGTAAACGGGAACATGCCCC 15527
QY 2108 TGCAGACA 2115
Db 15528 TGCAGACA 15535

RESULT 14
AE006464
LOCUS      AE006464      256073 bp      DNA
DEFINITION Homo sapiens 16p11.3 sequence section 3 of 8.
ACCESSION  AE006464 AE005175
VERSION     AE006464.1 GI:14336700
KEYWORDS
SOURCE
ORGANISM   human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1. (bases 1 to 256073)
            Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,
            Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and
            Higgs,D.R.
            Sequence, structure and pathology of the fully annotated terminal 2
            Mb of the short arm of human chromosome 16
            Hum. Mol. Genet. 10 (4), 339-352 (2001)
            1157797
JOURNAL    PUBMED
AUTHORS    Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,
            Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and
            Higgs,D.R.
TITLE      Direct Submission
JOURNAL    Submitted (07-DEC-2000) MRC Molecular Haematology Unit, Weatherall
            Institute of Molecular Medicine, John Radcliffe Hospital, Oxford,
            Oxon OX3 9DS, UK
FEATURES
    source
        1..256073
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /map="16p13.3"
            <1..20
                /note="GC-rich: RepeatMasker predicted
                23/10/2000"
            <1..3757
                /note="HS356B8"
            <1..4641
                /organism="Homo sapiens"
```







\* 133960 134059: gap of unknown length  
134060 155770: contig of 21711 bp in length.  
FEATURES  
source  
1. 155770  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-728H8"  
/clone.lib="RPCI human BAC library 11"  
BASE COUNT 29940 a 45687 c 45637 g 32240 t 2266 others.  
ORIGIN

Query Match 13.9% Score 296.6; DB 2; Length 155770;  
Best Local Similarity 82.2%; Pred. No. 1e-43; Indels 83; Gaps 2;  
Matches 402; Conservative 0; Mismatches 4;

QY 1710 GCTGCGTTGACTGTGCTGAGTGACAGAACTCTGGACCCACGTGAGAGGCTGAA 1769  
|||  
Db 21540 GCAGCGCTTGACTGTGCTGAGTGACAGAACTCTGGACCCACGTGAGAGGCTGAA 21481  
QY 1770 GCGGAGAGGAGGAGCCGCCGCG-GTGGGGGACTGATCTAGCGAGGCGAGGAGGAGC 1828  
|||  
Db 21480 GCGGAGAGGAGGAGCCGCCGCGGAGTGTGATCTAGCGAGGCGAGGAGGAGC 21421  
QY 1829 TGGACACGCTGGGGCTGGGGCTACAGGGCGGATCCCAACGGCTACCTGTCTAGAC 1888  
|||  
Db 21420 TGGACACGCTGGGGCTGGGGCTACAGGGCGGATCCCAACGGCTACCTGTCTAGAC 21361  
QY 1889 TCAGCGTCA----- 1898  
|||  
Db 21360 TCAGCATGCAAGGTGGGGGCGGCCAGGCCAGGGCTGGGGGCAAGCTGGGGGCTGG 21301  
QY 1899 -----AGAGACCTCTGGGGGAGCGGCTGCGCTC 1926  
|||  
Db 21300 AGGTGGGCGCTCTGAGTACCCCTCTCTGTAGAGGCCCTCTCGGGGACGCGCTGCTC 21241  
QY 1927 CTAGGACCTGAGACCTGTTCTCACCCTCTGGACACTGTCTAGCCTCCACCCCTGGCTGA 1986  
|||  
Db 21240 CTAGGACCTGAGACCTGTTCTCACCCTCTGGACACTGTCTAGCCTCCACCCCTGGCTGA 21181  
QY 1987 GGGCCCCACTCTCTGCTGGGCCCAAGCCCTGCTGGGATCCCGGCTGGCCAGAGCAGG 2046  
|||  
Db 21180 GGGCCCCACTCTCTGCTGGGCCCAAGCCCTGCTGGGATCCCGGCTGGCCAGAGCAGG 21121  
QY 2047 CACGGGTATCCCGTTCACCCCAAGAGAACTGCGCTCAGTAAAGGGAACATGCCCC 2106  
|||  
Db 21120 CACGGGTATCCCGTTCACCCCAAGAGAACTGCGCTCAGTAAAGGGAACATGCCCC 21061  
QY 2107 CTGCAGACA 2115  
|||  
Db 21060 CTGCAGACA 21052

Search completed: December 7, 2001, 11:24:02  
Job time: 4939 sec







[illegible]

Oy	1739	AAC TTC TGGA GAC CCCCACGCTGGAGGGCTTAAGCGGAGGAGCGGCACGCCGCCGGTGCCGG	1739
Db	1718	aactctcttggaacccccaagcttgagagccttgaaagcgtgaagagcggaccgcgccgtgtgcgag	1777
Oy	1799	ACTGATCTCTACGGCAGCAGCGGACAGACACTGGACACGGCTGGGGGCTGGGGGCTACAGGGCG	1858
Db	1778	acttgatctctacgcgcagcgcggacgaagacactctgaaacgcgtctggggctcggggcttaacgggcg	1837
Oy	1859	GCATCCCCCAACGGCTACCTGGTCTCTAGACCTCAGCGTGCAGAGAGACCCCTCTGGGGAGCG	1918
Db	1838	gcattccccaacggtctactcgtctctagactcaagcttgcaagagggccctctcgggagcgc	1897
Oy	1919	CCTGCTCTCTAGACCTGAGACCTGTCTACACGTCCTGTGGACATGCTCTCTACGCTCCACCC	1978
Db	1898	ctctgcctctctaggaaacctggaactcgtcttcaacgcttccctggacactgctctctagctctcaacc	1957
Oy	1979	TGGCGCTGAGGGCCCCACACTCCCTTGCTGCGGCCCGACGCGCTGCTGGGGATCCCCCGCTGGCA	2038
Db	1958	tggtctctgagggcccccaactctctcgtcgtgcgcctcgaagccctcgtcgggataccccgcctggcca	2017
Oy	2039	GGAGCAGGCACAGGGTGATCCCGCTTCCACCCCAAGAGAACTCGCGCTCAGTAAACGGGAA	2098
Db	2018	ggagcaggcacaggtgtgtcccgcttccaccccaaggaactcgcgtctagtaaacgggaa	2077
Oy	2099	CATGCCCCCTCTGACACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA	2138
Db	2078	catgccccctctgcagacacgtataaaaaaagaaaaaagaaaaaagaaaaaagaaaaaagaaaaaag	2117

RESULT	3	
AAAA64366		
ID	AAAA64366	standard; cDNA; 1207 BP.
XX		
AC	AAAA64366;	
XX		
DT	20-DEC-2000	(first entry)
XX		
DE	Soluble mesothelin related	(SMR) antigen cDNA sequence.
XX		
KW	Mesothelin; mesothelin related antigen; MRA; malignant condition; cancer	
KW	differentiation antigen; adenocarcinoma; mesothelioma; ovarian carcinoma;	
KW	pancreatic carcinoma; non-small cell lung carcinoma; MRA-2; SMR antigen;	
KW	ss.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	1..1206
FT		/*tag= a
FT		/product= "soluble mesothelin related (SMR) antigen"
FT		/note= "no termination codon given"
XX		
PN	WO200050900-A2.	
XX		
PD	31-AUG-2000.	
XX		
PF	25-FEB-2000; 2000WO-US04834.	
XX		
PR	26-FEB-1999; 99US-0121767.	
PR	05-AUG-1999; 99US-0147404.	
XX		
PA	(PACI-) PACIFIC NORTHWEST RES FOUND.	
XX		
PI	Scholler NB, Hellstrom I, Hellstrom KE;	
XX		
DR	WPI; 2000-572118/53.	
DR	P-PSDB; AAB08547.	
XX		
PT	Diagnosing carcinomas, e.g. adenocarcinoma or ovarian carcinomas, by	
PT	detecting reactivity of a molecule (with an antigenic determinant and	
PT	present in a sample) with an antibody specific for a mesothelin related	
PT	antigen polypeptide	
XX		







XX Example 10; Fig 6A-B; 92pp; English.

PS The present sequence encodes a human mesothelin related antigen (MRA)-2.  
 CC Antibodies specific to MRA polypeptides are used for the detection of a  
 CC malignant condition. Mesothelin is a differentiation antigen which is  
 CC expressed in the surfaces of normal mesothelial cells and also on  
 CC certain cancer cells, including epithelial ovarian tumours and  
 CC mesotheliomas. The method is useful for detecting the presence of a  
 CC malignant condition, specifically, adenocarcinoma, mesothelioma,  
 CC ovarian carcinoma, pancreatic carcinoma or non-small cell lung carcinoma.  
 XX  
 SQ Sequence 985 BP; 202 A; 295 C; 318 G; 170 T; 0 other;

Query Match 40.88; Score 871.4; DB 21; Length 985;  
 Best Local Similarity 97.38; Pred. No. 1.8e-163;  
 Matches 906; Conservative 0; Mismatches 1; Indels 24; Gaps 1;

QY 970 TTCGGCGGAGTGGAGAGACAGCTGTCTCTCAGCAAGAGGCGCGGAGATGAC 1029  
 Db 1 ttccgagcggaagtggagagacagctgtctctcaggaagagcccgagatagac 60  
 QY 1030 GAGAGCTCATCTTCTACAGAACTGGAGCTGGAGCCTGCTGATGCGGCCCTGCTG 1089  
 Db 61 gagagctcatctcttaacaagaatggagagctggagagctgagatgagccctgctg 120  
 QY 1090 GCCACCCAGATGGACCGCGTGAAGCCATCCCTTCACATCGAGCGAGCTGGAGCTGCTA 1149  
 Db 121 gccaccagatgagccgctgaacgcatccctcactcagagctgagagctgacta 180  
 QY 1150 AAGCATAACTGATGATGAGCTCTACCCAAAGGTTACCCCGAGTGTGATCCAGCACCTG 1209  
 Db 181 aagcataactgtagtgagctctaccacaaggtaccgccgagctgtgaltccagacctg 240  
 QY 1210 GGTACTCTTCTCTCAAGATGAGCCCTGAGGACATTCGCAAGTGGAATGTGACCTCCTG 1269  
 Db 241 ggtactctctctccaagatgagccctgagacatcgcgaatggaatgtagctcctg 300  
 QY 1270 GAGACCGCTGAAGGCTTGTGATGAGCAAGGCAAGCAAGATCTCTCAGAGCTCCT 1329  
 Db 301 gagaccctgagagctgtctgtgaagtcacaagaaggacgaatgagtcct----- 351  
 QY 1330 CGGCGGCCCTCCACAGATGGCCACCTGATCGACCGCTTGTGAAGGGAAGGGCGAG 1389  
 Db 352 -----caggggccaacctgacgcgcttggaaagggaaggggcag 396  
 QY 1390 CTAGACAAAGACACCTTAGACACCTGACCGCTTCTACCTGGGTACTGTGCTCCCTC 1449  
 Db 397 ctgagacaaagacacctagacacctgacgcctctacacctgggtactgtgtcctc 456  
 QY 1450 AGCCCGAGAGAGCTGAGCTCGGTGCCCCCGACAGCATCTGGGGGGTCAAGGCCCCAGAG 1509  
 Db 457 agcccgagagagctgagctccggtgcccccgacgacatcggggtcgaagccccagag 516  
 QY 1510 CTGGACACGTGTGACCAAGGACAGCTGAGCTCTCTATCCCAAGGCGCGCTTGTCTTC 1569  
 Db 517 ctggacacggtgtgacccaagagcagctgagctctctatcccaagggcgcttgccttc 576  
 QY 1570 CAGAACATGAACGGGTGCGAATACTTGTGAAGATCCAGTCTTCTGGTGGGCCCCC 1629  
 Db 577 cagaaactgaacgggttcgaaactctgtgaagatccagtcctccgggtggggcccc 636  
 QY 1630 ACGGAGATTGGAAGGCGCTCAGTCAGACAGATGAGCATGAGCATGAGCATGAGCATGATG 1689  
 Db 637 acggagatttggaaggcgctcagtcagagaatgtgagcatgagcttggccaagctcatg 696  
 QY 1690 AAGCTGGGACGAGTGGGTGCTGCGCTTGAAGTGTGCTGAGTGCAGAAATTTCTGGGA 1749  
 Db 697 aagctgggacgagtggtgctgctgctgactgtgagtgagtgagaaactcttgga 756  
 QY 1750 CCCCACTGAGGGGCTGGAAGGCGAGGAGCGGCGCGCGGCTGGGAGCTGATCCTA 1809  
 Db 1750 ccccaactgaggggctggaaggcgagagagcgagcagcccgctggcgagctgataccta 816

Db 757 ccccaactgaggggctggaaggcgagagagcgagcagcccgctggcgagctgataccta 816  
 QY 1810 CGGACGCGGACAGACACCTGAGACACCTGAGGCTGGGCTACAGGGCGGATCCCCAAC 1869  
 Db 817 cggacgcgagcagagacactgagacagctgggtctgggtctacagggcgatccccaac 876  
 QY 1870 GGTACCGTGGTCTCAGACCTCAGCGTGCAG 1900  
 Db 877 ggtaccgtggctcctagacctcagctgcaag 907

# RESULT 6

AAFP94935  
 ID AAFP94935 standard; cDNA; 396 BP.

AC AAFP94935;  
 DT 23-MAY-2001 (first entry)  
 XX

DE Human ovarian cancer associated coding sequence SEQ ID NO: 126;  
 XX

KM Human, ovarian cancer; vaccine; gene therapy; carcinoma; ss.  
 XX

OS Homo sapiens.  
 XX

PN W0200118046-A2.  
 XX

PD 15-MAR-2001.  
 XX

PF 08-SEP-2000; 2000MO-US24827.  
 XX

PR 10-SEP-1999; 99US-0394374.  
 XX

PR 01-MAY-2000; 2000US-0561778.  
 XX

PR 15-AUG-2000; 2000US-0640173.  
 XX

PR 07-SEP-2000; 2000US-0656668.  
 XX

PA (CORI-) CORIXA CORP.  
 XX

PI Xu J, Stolk JA.  
 XX

DR WPI; 2001-211395/21.  
 XX

XX Isolated polypeptides associated with ovarian carcinomas, and the  
 PT nucleic acids that encode them, useful for the prevention diagnosis and  
 PT treatment of ovarian cancers -  
 XX

PS Claim 18; Page 157; 189pp; English.  
 XX

XX The present invention provides a number of coding sequences and proteins,  
 CC the over-expression of which is associated with ovarian carcinoma/cancer.  
 CC These can be used in the diagnosis, treatment and prevention of ovarian  
 CC cancer, optionally by gene therapy or in the form of a vaccine. The  
 CC present sequence is an example of one of these sequences.  
 XX

XX Sequence 396 BP; 77 A; 139 C; 111 G; 69 T; 0 other;  
 SQ

Query Match 10.38; Score 219.8; DB 22; Length 396;  
 Best Local Similarity 80.98; Pred. No. 9.3e-35;  
 Matches 279; Conservative 0; Mismatches 42; Indels 24; Gaps 1;

QY 1243 ATTGCAAGTGAATGTGACGTCCCTGAGACACCTGAAGCCTTGTGATGAGCAAA 1302  
 Db 1 attgcaagtgaatgtgacgtccctgagacacctgaagccttgtgatgagcaaaa 67

QY 1303 GGGACGAATGACTCCTCAGGCTCTTCGCGGCCCTCCACAGGTGGCCACCTGATC 1362  
 Db 68 gggacgaatgactcctcaggctcttcgcggccctccacaggtggccacctgatic 103

QY 1363 GACCGCTTGTGAAGGGAAGGCGCAGCTAGACAAACACCCAGACACCTGACCGCC 1422  
 Db 104 gaccgcttgtgaaagggaaggcgagcagctagaacaaagacacctagacacccgacgc 163

QY	1423	TCACCCCTGGGATACGTCGTCCTCCCTCAGCCCCGAGAGACTGAGTTCGTGGCCCCAGC	1482
Db	164	ttcacccctggtactcgtgctctccctcagccccgagactgtagctccgtgcgcccccagc	223
QY	1483	AGCATCTGGGGCGGTCAAGGCCCCAGAGACTGTGACAGCTGTGACCCCAAGGACGTGGACGTC	1542
Db	224	agcatctggggcggcgaagcccccaagcctcggaacgctggtggctcaggtctcaagatgagcggc	283
QY	1543	CTGTATCCCAAGGCCCGCTTGTCTTTCAGAAACATGATACGGGTCC	1587
Db	284	atcccccaagcgctactcgtgctctcagaccctcagatcgtcaagaagggc	328

Result	7	
AA063971		
ID	AA063971	standard; cDNA; 197 BP.
XX		
AC	AA063971;	
XX		
DT	08-DEC-1994	(first entry)
XX		
DE	Meg-Pot fragment.	
XX		
KW	Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;	
RV	platelet; amplification; primer; polymerase chain reaction; PCR; ss	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	3..197
FT		/*tag= a
FT		/note= "Meg-Pot fragment"
XX		
PN	W09410312-A.	
XX		
PD	11-MAY-1994.	
XX		
PF	25-OCT-1993;	93WO-JP01540.
XX		
PR	23-OCT-1992;	92JP-0286153.
PR	11-NOV-1992;	92JP-0301387.
PR	09-DEC-1992;	92JP-0329546.
XX		
PA	(CHUS ) CHUGAI SEIYAKU KK.	
XX		
PI	Hattori K, Kojima T, Oh-eda M, Yamaguchi N;	
XX		
DR	WPI; 1994-167467/20.	
XX		
DR	P-PSDB; AAR53991.	
XX		
PT	New megakaryocyte potentiator - for potential treatment of	
PT	thrombocytopenia	
XX		
PS	Disclosure; Page 51; 74pp; Japanese.	
XX		
CC	Use of the primers given in AAQ07809-10 in PCR resulted in the	
CC	fragments given in AA063971.	
CC	DNA encoding Meg-Pot has potential use in treatment of	
CC	thrombocytopenia and low platelet function.	
XX		
XX	Sequence 197 BP; 32 A; 66 C; 59 G; 40 T; 0 other;	

```

Query Match      9.2%; Score 197; DB 15; Length 197;
Best Local Similarity 100.0%; Pred. No. 2.7e-30;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

[illegible]

Db 61 tgcctcgtgacttcgacagagaatgacagctctcaacagagcagctgcgtcctcg 120  
Oy 407 CTACACGGCTCTCTGTAAGCCCCCGACAGACCTGAGACGGCTCCCTTGGACCTGGTAT 466  
Db 121 ctcaacggcctctctcgaagcccccgaagacctctgcagccctccacttgagactctgctat 180  
Oy 467 TCTCTACCCAGATGCG 483  
Db 181 tctcaacccagatgcg 197

RESULT 8  
 AAH81882  
 ID AAH81882 standard; DNA; 255 BP.  
 XX  
 AC AAH81882;  
 XX  
 DT 21-SEP-2001 (first entry)  
 XX  
 DE Rat differential transcription-associated cDNA SEQ ID 391.  
 XX  
 KW Differential transcription; human; rat; tumour cell; cytostatic;  
 KW Ras modulator; Class II tumour suppressor gene; gene therapy; ss.  
 XX  
 OS Rattus sp.  
 XX  
 PN WO200157058-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 31-JAN-2001; 2001WO-EP01003.  
 XX  
 PR 31-JAN-2000; 2000DE-1004102.  
 XX  
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX  
 PI Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;  
 PI Grips M, Hellriegel M, Schmitz A, Sers C;  
 XX  
 DR WPI; 2001-483415/52.  
 XX  
 PT Nucleic acids differentially expressed between tumor and normal cells,  
 PT useful for diagnosis or therapy of tumors and for screening active  
 PT agents -  
 XX  
 XX  
 PS Claim 6; Page 488; 579pp; German.  
 XX  
 CC This invention describes a nucleic acid (I) with differential expression  
 CC between tumour and normal cells and which has cytostatic activity. (I)  
 CC work as modulators of Ras actively by inducing expression of tumour  
 CC suppressor genes. (I), and polypeptides encoded by them, are useful as  
 CC targets for diagnosis or therapy and in screening to determine the  
 CC effects of an active compound (potential pharmaceutical) on a cell line,  
 CC particularly for diagnosis and treatment of tumors, especially by  
 CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme  
 CC methods) or by modulating the amount and/or location of (I)-encoded  
 CC polypeptides (by administration of the polypeptide or its activator,  
 CC antibody (optionally as a conjugate) or inhibitor). The method allows  
 CC identification of many Class II tumour suppressor genes (i.e. genes that  
 CC are not primary targets for tumour-initiating mutations).  
 CC AAH81492-AAH82376 represent the human and rat derived nucleic acid  
 CC fragments described in the method of the invention.  
 XX  
 Sequence 255 BP; 71 A; 71 C; 59 G; 54 T; 0 other;

Query Match	7.0%;	Score 150.2;	DB 22;	Length 255;
Best Local Similarity	-76.1%;	Pred. No. 4.9e-21;		
Matches 185;	Conservative 0;	Mismatches 58;	Indels 0;	Gaps 0;

QY 1082 CCCTGCTGGCCACCAGATGTGACCGGGTGAAGGCCATCCCCTTCACCTTAGAGAGACTGG 11411  
|||||  
Db 2 ccctgtcgtccgagccagatgatgccttgtaatgaattcccttacctacgacagactca 61



Query Match 4.2%; Score 89.8; DB 16; Length 95;  
 Best Local Similarity 95.8%; Pred. No. 3.6e-09;  
 Matches 91; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2023 GATCCCGCCCTGGCCAGGACGAGCGGTGATCCCGTTCACACCCAGAGAAGTCCG 2082  
 |||  
 Db 1 gatcccgccctggccagagcagcggtgctcccggttcacacccagagagactcgc 60  
 QY 2083 GCTCAGTAACGGGACATGCCCTCGACACAA 2117  
 |||  
 Db 61 gctcagtaaacgggacatgccctctgacgacaa 95

## RESULT 11

AA063970  
 ID AA063970 standard; cDNA; 75 BP.  
 AC AA063970;  
 XX  
 DT 08-DEC-1994 (first entry)  
 DE Meg-Pot fragment.  
 XX  
 KW Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;  
 KM platelet; amplification; primer; polymerase chain reaction; PCR; ss.  
 XX  
 PA Homo sapiens.  
 OS  
 PN W09410312-A.  
 PD 11-MAY-1994.  
 XX  
 PF 25-OCT-1993; 93MO-JP01540.  
 XX  
 PR 23-OCT-1992; 92JP-0286153.  
 PR 11-NOV-1992; 92JP-0301387.  
 PR 09-DEC-1992; 92JP-0329546.  
 XX  
 PA (CHUS ) CHUGAI SEIYAKU KK.  
 XX  
 PI Hattori K, Kojima T, Oh-eda M, Yamaguchi N;  
 XX  
 DR WPI: 1994-167467/20.  
 DR P-PSDB; AAR53990.  
 XX  
 PT New mega.karyocyte potentiator - for potential treatment of  
 PT thrombocytopenia  
 XX  
 PS Disclosure; Page 50; 74pp; Japanese.  
 XX  
 CC Use of the primers given in AAQ77805-08 in PCR resulted in the  
 CC fragments given in AA063969-70.  
 CC DNA encoding Meg-Pot has potential use in treatment of  
 CC thrombocytopenia and low platelet function.  
 CC  
 XX  
 SQ Sequence 75 BP; 11 A; 31 C; 16 G; 17 T; 0 other;

Query Match 3.5%; Score 75; DB 15; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-06;  
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 CTGCTGAGCCCGGAGAGCTGAGCGCTCCCATTTGAGCGCTATTCCTCAAC 474  
 |||  
 Db 1 ctctctgagcccgagagctgagcgctcccatctgagcgctatctcccaac 60  
 QY 475 CCAGATGCGTCTCG 489  
 |||  
 Db 61 ccagatgctctcgc 75

## RESULT 12

AA063969  
 ID AA063969 standard; cDNA; 69 BP.  
 AC AA063969;  
 XX  
 DT 08-DEC-1994 (first entry)  
 DE Meg-Pot fragment.  
 XX  
 KW Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;  
 KM platelet; amplification; primer; polymerase chain reaction; PCR; ss.  
 XX  
 PA Homo sapiens.  
 OS  
 PN W09410312-A.  
 PD 11-MAY-1994.  
 XX  
 PF 25-OCT-1993; 93MO-JP01540.  
 XX  
 PR 23-OCT-1992; 92JP-0286153.  
 PR 11-NOV-1992; 92JP-0301387.  
 PR 09-DEC-1992; 92JP-0329546.  
 XX  
 PA (CHUS ) CHUGAI SEIYAKU KK.  
 XX  
 PI Hattori K, Kojima T, Oh-eda M, Yamaguchi N;  
 XX  
 DR WPI: 1994-167467/20.  
 DR P-PSDB; AAR53989.  
 XX  
 PT New mega.karyocyte potentiator - for potential treatment of  
 PT thrombocytopenia  
 XX  
 PS Disclosure; Page 50; 74pp; Japanese.  
 XX  
 CC Use of the primers given in AAQ77805-08 in PCR resulted in the  
 CC fragments given in AA063969-70.  
 CC DNA encoding Meg-Pot has potential use in treatment of  
 CC thrombocytopenia and low platelet function.  
 CC  
 XX  
 SQ Sequence 69 BP; 7 A; 22 C; 26 G; 14 T; 0 other;

Query Match 3.2%; Score 69; DB 15; Length 69;  
 Best Local Similarity 100.0%; Pred. No. 4.4e-05;  
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 CGCCNACTCTTGCTTCCTCCGTGCGAGGTGTCGGCTGAGCAGGAGCGTCCGG 342  
 |||  
 Db 1 cgcacactcttgcttcctccgtgctcgaggtgctcgagcagcgagcgctgcgcg 60  
 QY 343 GAGCTGCGT 351  
 |||  
 Db 61 gagctgct 69

## RESULT 13

AA063969  
 ID AA063969 standard; cDNA to mRNA; 68 BP.  
 AC AA063969;  
 XX  
 DT 10-OCT-1996 (first entry)  
 DE Human gene signature HUMGS07964.  
 XX  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 XX  
 OS Homo sapiens.  
 XX







**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2001, 10:01:57 ; Search time 55.33 seconds  
(without alignments)  
8751.310 Million cell updates/sec

Title: US-09-215-035-1

Perfect score: 2138  
Sequence: 1 AGGATTCGGCGGCCGCC.....AAAAAAAAAAAAAAAAAAAA 2138

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11338999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/1na/PCTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/1na/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2138	100.0	2138	3	US-08-776-271-1 Sequence 1, Appl
2	2138	100.0	2138	3	US-09-215-035-1 Sequence 1, Appl
3	1994	93.3	2129	1	US-08-426-819A-34 Sequence 34, Appl
4	197	9.2	197	1	US-08-426-819A-32 Sequence 32, Appl
5	75	3.5	75	1	US-08-426-819A-31 Sequence 31, Appl
6	69	3.2	69	1	US-08-426-819A-30 Sequence 30, Appl
7	52.4	2.5	3407	1	US-08-253-155A-7 Sequence 7, Appl
8	46.4	2.2	2580	3	US-09-050-863-2 Sequence 2, Appl
9	46.4	2.2	5452	2	US-09-130-114-1 Sequence 1, Appl
10	46.4	2.2	9600	4	US-08-910-647-1 Sequence 1, Appl
11	46.4	2.2	10596	1	US-07-885-871-15 Sequence 15, Appl
12	46.4	2.2	10596	1	US-08-087-783A-15 Sequence 15, Appl
13	46.4	2.2	10596	1	US-08-087-783A-15 Sequence 15, Appl
14	46.4	2.2	10596	1	US-08-194-088B-15 Sequence 15, Appl
15	46.4	2.2	10596	2	US-08-194-087-15 Sequence 15, Appl
16	46.4	2.2	10596	5	PCT-US93-04648-15 Sequence 15, Appl
17	46.4	2.2	71989	4	US-09-443-501A-2 Sequence 2, Appl
18	45.2	2.1	1505	1	US-07-915-246-1 Sequence 1, Appl
19	44.6	2.1	4257	2	US-08-690-473-1 Sequence 1, Appl
20	44.6	2.1	4257	4	US-09-259-821A-1 Sequence 1, Appl
21	44.6	2.1	4257	4	US-08-843-659-1 Sequence 1, Appl
22	44.6	2.1	12001	1	US-08-458-568A-11 Sequence 11, Appl
23	44	2.1	1771	2	US-08-533-669A-7 Sequence 7, Appl
24	44	2.1	1771	2	US-08-511-872-1 Sequence 1, Appl
25	43.2	2.0	20235	1	US-07-642-734C-3 Sequence 3, Appl
26	43.2	2.0	20235	1	US-08-439-009A-3 Sequence 3, Appl
27	43	2.0	1690	3	US-08-869-696-8 Sequence 8, Appl

28	42.6	2.0	4524	2	US-08-845-998-7 Sequence 7, Appl
29	42.6	2.0	4524	3	US-09-206-537-7 Sequence 7, Appl
30	42.6	2.0	4524	4	US-09-430-854-7 Sequence 21, Appl
31	42.4	2.0	1282	3	US-08-869-696-21 Sequence 21, Appl
32	42.4	2.0	1372	3	US-08-869-696-22 Sequence 22, Appl
33	42.4	2.0	5529	3	US-08-869-696-1 Sequence 1, Appl
34	41.8	2.0	31571	1	US-08-323-443B-1 Sequence 1, Appl
35	41.6	1.9	4411529	4	US-09-103-840A-1 Sequence 1, Appl
36	41.4	1.9	1931	2	US-05-130-114-2 Sequence 2, Appl
37	41.4	1.9	2214	3	US-08-864-038A-1 Sequence 1, Appl
38	41.4	1.9	3331	3	US-08-864-038A-2 Sequence 2, Appl
39	41.4	1.9	3331	3	US-08-864-038A-4 Sequence 4, Appl
40	41.2	1.9	1995	1	US-08-425-069-3 Sequence 3, Appl
41	41.2	1.9	1995	2	US-08-317-844B-3 Sequence 3, Appl
42	41.2	1.9	2793	1	US-08-209-747-1 Sequence 1, Appl
43	41.2	1.9	2793	1	US-08-458-298-1 Sequence 1, Appl
44	41	1.9	35100	2	US-08-770-379-17 Sequence 17, Appl
45	41	1.9	35100	4	US-08-757-669A-17 Sequence 17, Appl

#### ALIGNMENTS

RESULT 1  
US-08-776-271-1  
Sequence 1, Application US/08776271  
Patent No. 6083502  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
TITLE OF INVENTION: Mesothelin, a Differentiation Antigen  
TITLE OF INVENTION: Present on Mesothelium, Mesotheliomas and Ovarian Cancers  
TITLE OF INVENTION: and Methods and Kits for Targeting the Antigen  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,271  
FILING DATE: 01-DEC-1998  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/00224  
FILING DATE: 03-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/010,166  
FILING DATE: 05-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Paris, Susan K. 41,739  
REGISTRATION NUMBER: 41,739  
REFERENCE/DOCKET NUMBER: 015280-259100US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2138 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 100..1986

US-08-776-271-1

Query Match 100.0%; Score 2138; DB 3; Length 2138;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAATTCGGGTGGGCGGACCTCCGCTGTGTGACGGCGGACAGAGACTACCGGT 60  
DB 1 AGGAATTCGGGTGGGCGGACCTCCGCTGTGTGACGGCGGACAGAGACTACCGGT 60

QY 61 GGACCCACGGGTGCTCCCTCCCTGGGATCTACACAGACATGGCTTGCAAGGCTGAC 120  
DB 61 GGACCCACGGGTGCTCCCTCCCTGGGATCTACACAGACATGGCTTGCAAGGCTGAC 120

QY 121 CCCGTTGGTCTGTGGGGACGGCCGTGGACGCTCCGTTCCGCTGCTTACGCTTGA 180  
DB 121 CCCGTTGGTCTGTGGGGACGGCCGTGGACGCTCCGTTCCGCTGCTTACGCTTGA 180

QY 181 TGGGTGCTATCCCGGAGAGACCTGGCTGGAGAGACAGGGAGAGTGCCTGGGG 240  
DB 181 TGGGTGCTATCCCGGAGAGACCTGGCTGGAGAGACAGGGAGAGTGCCTGGGG 240

QY 241 GGAATCTGACAAACCCCATTAACATTTCCAGCTCTCCCTCCGCAACTCTTGCTTC 300  
DB 241 GGAATCTGACAAACCCCATTAACATTTCCAGCTCTCCCTCCGCAACTCTTGCTTC 300

QY 301 CCGTGTGGGAGGTGTCCGGGCTGAGCAGGAGGCTGTCCGGAGAGTGGTGTGGGCTTG 360  
DB 301 CCGTGTGGGAGGTGTCCGGGCTGAGCAGGAGGCTGTCCGGAGAGTGGTGTGGGCTTG 360

QY 361 GCACAGAAAGATGTCAAGCTCTCAACAGAGAGAGCTGGCTGTGGCTCACCGGCTCT 420  
DB 361 GCACAGAAAGATGTCAAGCTCTCAACAGAGAGAGCTGGCTGTGGCTCACCGGCTCT 420

QY 421 GAGCCCCCGGAGACCTTGAGAGCCCTCCCATTTGAACTGCTGTATTCTTAACCCAGAT 480  
DB 421 GAGCCCCCGGAGACCTTGAGAGCCCTCCCATTTGAACTGCTGTATTCTTAACCCAGAT 480

QY 481 GCGTTCTGGGGGCGCCAGGAGCTGCACCCGTTTCTCCGCTATCAGAGGCAATGTG 540  
DB 481 GCGTTCTGGGGGCGCCAGGAGCTGCACCCGTTTCTCCGCTATCAGAGGCAATGTG 540

QY 541 GACCTGTCTCCGAGAGGGGGCTCCGAGAGAGAGAGGAGCTGTGGCTGTGGCTGTG 600  
DB 541 GACCTGTCTCCGAGAGGGGGCTCCGAGAGAGAGAGGAGCTGTGGCTGTGGCTGTG 600

QY 601 TGGGCTGTGGGGGGGTCTCTGCTGAGAGAGCTGATGTGGGGCTTGAGAGGCTTGCT 660  
DB 601 TGGGCTGTGGGGGGGTCTCTGCTGAGAGAGCTGATGTGGGGCTTGAGAGGCTTGCT 660

QY 661 TGGCAGCTGCTGGGGGCTTTGTGGCCGAGTGGCCGAAAGTGTCTACCCGGCTGTG 720  
DB 661 TGGCAGCTGCTGGGGGCTTTGTGGCCGAGTGGCCGAAAGTGTCTACCCGGCTGTG 720

QY 721 AGCTGTCCCGGAGACCTTGAGACAGAGACAGAGAGAGGAGGCGGCTCTGCAAGGC 780  
DB 721 AGCTGTCCCGGAGACCTTGAGACAGAGACAGAGAGAGGAGGCGGCTCTGCAAGGC 780

QY 781 GGGGAGACCCCTTACGGGCCCCCGCTGACATGTGTCTCTCAGAGTGAAGCTTGCGG 840  
DB 781 GGGGAGACCCCTTACGGGCCCCCGCTGACATGTGTCTCTCAGAGTGAAGCTTGCGG 840

QY 841 GGGCGTGTGGGCTGTGGGGGAGGCGCATCATCGGACATCTCCGAGAGGATGCTGGCC 900  
DB 841 GGGCGTGTGGGCTGTGGGGGAGGCGCATCATCGGACATCTCCGAGAGGATGCTGGCC 900

QY 901 GCGTGTGGGCGCAACGCTCTCTCTGGGAGCCCATCTGCGGGAGCTTGAAGGAGCATCTCTC 960  
DB 901 GCGTGTGGGCGCAACGCTCTCTCTGGGAGCCCATCTGCGGGAGCTTGAAGGAGCATCTCTC 960

QY 961 CGGCGGCGGTTCCGGGAGGAGTGAAGACAGCTGTCTCTTACGAGCAAGAGGCGCCG 1020  
DB 961 CGGCGGCGGTTCCGGGAGGAGTGAAGACAGCTGTCTCTTACGAGCAAGAGGCGCCG 1020

DB 961 CGGCGGCGGTTCCGGGAGGAGTGAAGACAGCTGTCTCTTACGAGCAAGAGGCGCCG 1020

QY 1021 GAGATGACGAGAGCTCATCTTCTACAGAGAGTGGAGCTGGAAGCTGCTGGATGCG 1080  
DB 1021 GAGATGACGAGAGCTCATCTTCTACAGAGAGTGGAGCTGGAAGCTGCTGGATGCG 1080

QY 1081 GCCCTGTGGCCACCCAGATGAGACCGGCTGAAGCCATCCCTTTCACCTACGACACCTG 1140  
DB 1081 GCCCTGTGGCCACCCAGATGAGACCGGCTGAAGCCATCCCTTTCACCTACGACACCTG 1140

QY 1141 GAGCTCTTAAAGCATTAACCTGATGAGCTCTACCCACAAGTTACCCCGATCTGTATC 1200  
DB 1141 GAGCTCTTAAAGCATTAACCTGATGAGCTCTACCCACAAGTTACCCCGATCTGTATC 1200

QY 1201 CAGACCTGGGCTACCTCTTCTCAAGATGAGCCCTGAGACATTCGCAAGTGAATGTG 1260  
DB 1201 CAGACCTGGGCTACCTCTTCTCAAGATGAGCCCTGAGACATTCGCAAGTGAATGTG 1260

QY 1261 ACGTCCCTGAGACCTTGAAAGCTTGTGTAAGTGCACAAAGGGGCAAGAAATGATCT 1320  
DB 1261 ACGTCCCTGAGACCTTGAAAGCTTGTGTAAGTGCACAAAGGGGCAAGAAATGATCT 1320

QY 1321 CAGGCTCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1380  
DB 1321 CAGGCTCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1380

QY 1381 AGGGGCGAGCTAGACAAAGACACCTAGACACCTGACCGCTTCTACCTTGGTACTG 1440  
DB 1381 AGGGGCGAGCTAGACAAAGACACCTAGACACCTGACCGCTTCTACCTTGGTACTG 1440

QY 1441 TGTCTCTCAGCCCGGAGAGCTAGCTCGTGCCTCCCGACAGCTATCTGGGCGGTAGG 1500  
DB 1441 TGTCTCTCAGCCCGGAGAGCTAGCTCGTGCCTCCCGACAGCTATCTGGGCGGTAGG 1500

QY 1501 CCCCAGAGACCTGAGACGCTGTGACCCAGAGGAGCTGAGACGCTCTTATCCCAAGGCCG 1560  
DB 1501 CCCCAGAGACCTGAGACGCTGTGACCCAGAGGAGCTGAGACGCTCTTATCCCAAGGCCG 1560

QY 1561 CTTGCTTTTCCAGAACATGAAAGGGGTCCGAATCTTCTGGAAGATCCAGTCTTCTGGGT 1620  
DB 1561 CTTGCTTTTCCAGAACATGAAAGGGGTCCGAATCTTCTGGAAGATCCAGTCTTCTGGGT 1620

QY 1621 GGGGCCCCCAGAGAGATTTGAAGGCGCTCAGTCAGAGATGTGAGCATGATGAGCTTGGCC 1680  
DB 1621 GGGGCCCCCAGAGAGATTTGAAGGCGCTCAGTCAGAGATGTGAGCATGATGAGCTTGGCC 1680

QY 1681 ACGTTCAATGAAGCTGGGAGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740  
DB 1681 ACGTTCAATGAAGCTGGGAGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740

QY 1741 CTTCTGGAGCCCAAGTGAAGGGCTGGAAGGCGGAGAGGCGCACCCGCGGTGCGGAGAC 1800  
DB 1741 CTTCTGGAGCCCAAGTGAAGGGCTGGAAGGCGGAGAGGCGGAGAGGCGCGGTGCGGAGAC 1800

QY 1801 TGGATCTTACGGGAGCGGAGAGAGCTGGAACAGCTGGGGCTTGAGAGGCGGC 1860  
DB 1801 TGGATCTTACGGGAGCGGAGAGAGCTGGAACAGCTGGGGCTTGAGAGGCGGC 1860

QY 1861 ATCCCAAGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920  
DB 1861 ATCCCAAGGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920

QY 1921 TGGCTCTTACGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980  
DB 1921 TGGCTCTTACGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980

QY 1981 GCGTGAAGGGGCGCACTCTCTTGTGGCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040  
DB 1981 GCGTGAAGGGGCGCACTCTCTTGTGGCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040

QY 2041 AGCAGGACAGGGGTGATCCCGCTTCCACCCCAAGAGAACTCCGCTCAGTAAGGAGGAA 2100  
DB 2041 AGCAGGACAGGGGTGATCCCGCTTCCACCCCAAGAGAACTCCGCTCAGTAAGGAGGAA 2100

OY 2101 TGCCCCCTGCAGCAAAAAAAAAAAAAAAAAAAAA 2138  
Db 2101 TGCCCCCTGCAGCAAAAAAAAAAAAAAAAAAAAA 2138

RESULT 2  
US-09-215-035-1  
Sequence 1, Application US/09215035  
Patent No. 6153430  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Chang, Kai  
TITLE OF INVENTION: Mesothelin, a Differentiation Antigen  
TITLE OF INVENTION: Present on Mesothelium, Mesotheliomas and Ovarian Cancers  
TITLE OF INVENTION: and Methods and Kits for Targeting the Antigen  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
City: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/215,035  
FILING DATE: No. 6153430 yet assigned  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/776,271  
FILING DATE: 01-DEC-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/00224  
FILING DATE: 03-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/010,166  
FILING DATE: 05-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Paris, Susan K.  
REGISTRATION NUMBER: 41,739  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2138 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 100..1986  
US-09-215-035-1

Query Match 100.0% Score 2138: DB 3: Length 2138:  
Best Local Similarity 100.0%: Pred. No.0:  
Matches 2138: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 AGGAATTCGCGTGGCGCGGCACTCCCTGCTGCTGACGCGCGAGAGAGTACCGGT 60  
Db 1 AGGAATTCGCGTGGCGCGGCACTCCCTGCTGCTGACGCGCGAGAGAGTACCGGT 60  
OY 61 GGACCAAGGCTGCTCCCTCCCTGGGATCTACACAGACCATGCGCTTGGCAAGGCTCGAC 120  
Db 61 GGACCAAGGCTGCTCCCTCCCTGGGATCTACACAGACCATGCGCTTGGCAAGGCTCGAC 120

OY 121 CCGTGTGTCCTGCTGGGGAGCCGCTGGCAGCCTCCGTTCTGCTCTTACGCTCGGA 180  
Db 121 CCGTGTGTCCTGCTGGGGAGCCGCTGGCAGCCTCCGTTCTGCTCTTACGCTCGGA 180  
OY 181 TGCGTGCATCCCGGAGAGACCTTGCTGAGAGACAGAGAGAGTGTGCCCCCTGGGG 240  
Db 181 TGCGTGCATCCCGGAGAGACCTTGCTGAGAGACAGAGAGAGTGTGCCCCCTGGGG 240  
OY 241 GGAATCTGACACACCCCATACATTTCACGCTCTCCCTGGCAACTCTTGCTTC 300  
Db 241 GGAATCTGACACACCCCATACATTTCACGCTCTCCCTGGCAACTCTTGCTTC 300  
OY 301 CCGGTGCGGAGGTGTGCGGCTGAGCAGAGAGGTGTCGGGAGAGTGGGTGGGCTTG 360  
Db 301 CCGGTGCGGAGGTGTGCGGCTGAGCAGAGAGGTGTCGGGAGAGTGGGTGGGCTTG 360  
OY 361 GCACAGAAGATGTCAAGCTCTCAACAGACAGCTGCGCTGTGAGCTCAACCGCTCT 420  
Db 361 GCACAGAAGATGTCAAGCTCTCAACAGACAGCTGCGCTGTGAGCTCAACCGCTCT 420  
OY 421 GAGCCCCCGAGACCTGAGACGCTCCCATTTGACCTGCTGTATTCTACCCAGAT 480  
Db 421 GAGCCCCCGAGACCTGAGACGCTCCCATTTGACCTGCTGTATTCTACCCAGAT 480  
OY 481 GCGTTCGCGGGCCCGAGGCTGACCCGTTCTCTCCGACATCAGAAAGCCAAATGTG 540  
Db 481 GCGTTCGCGGGCCCGAGGCTGACCCGTTCTCTCCGACATCAGAAAGCCAAATGTG 540  
OY 541 GACCTGCTCCCGAGAGGGGGCTCCGAGCAGACAGGCGTGTGCTGCGGCTGCGCTGC 600  
Db 541 GACCTGCTCCCGAGAGGGGGCTCCGAGCAGACAGGCGTGTGCTGCGGCTGCGCTGC 600  
OY 601 TGCGGTGTGCGGGGGTCTGTCTGAGCAGGCTGATGTGCGGGCTGTGGAGGCTGGCT 660  
Db 601 TGCGGTGTGCGGGGGTCTGTCTGAGCAGGCTGATGTGCGGGCTGTGGAGGCTGGCT 660  
OY 661 TGCGACGTGCGGGGGCTTTGTGCGGAGTGTGCTGAGGAGGCTGTGGAGGCTGGCT 720  
Db 661 TGCGACGTGCGGGGGCTTTGTGCGGAGTGTGCTGAGGAGGCTGTGGAGGCTGGCT 720  
OY 721 AGCTGCCCGGAGCCCTGTGACAGCAGACAGAGGAGGAGGCGGCTGTGAGGCG 780  
Db 721 AGCTGCCCGGAGCCCTGTGACAGCAGACAGAGGAGGAGGCGGCTGTGAGGCG 780  
OY 781 GGGGAGCCCCCTACGGCCCCCGCTGACATGTCTCTCAGAGATGAGAGCTGTGGG 840  
Db 781 GGGGAGCCCCCTACGGCCCCCGCTGACATGTCTCTCAGAGATGAGAGCTGTGGG 840  
OY 841 GGCCTGCGCCGCTGTGCGGCGACCCCATCTCCGACAGATCCCGAGGCGATGCTGCG 900  
Db 841 GGCCTGCGCCGCTGTGCGGCGACCCCATCTCCGACAGATCCCGAGGCGATGCTGCG 900  
OY 901 GCGTGGGCGCAACGCTCTCTGCGGAGCCATCTGCGGCGAGCTTGAAGGACCATCTC 960  
Db 901 GCGTGGGCGCAACGCTCTCTGCGGAGCCATCTGCGGCGAGCTTGAAGGACCATCTC 960  
OY 961 CGCGCGGGGTTCGCGGAGAGTGGAGAACGACGCTGTCTTACAGGCAAAAGGCCCGC 1020  
Db 961 CGCGCGGGGTTCGCGGAGAGTGGAGAACGACGCTGTCTTACAGGCAAAAGGCCCGC 1020  
OY 1021 GAGATAGACGAGACCTCATCTTCTACAAAGTGGAGCTGGAGGCTGGTGGATGCG 1080  
Db 1021 GAGATAGACGAGACCTCATCTTCTACAAAGTGGAGCTGGAGGCTGGTGGATGCG 1080  
OY 1081 GCGCTGTGCGCCACCAAGATGAGCCGCTGAAGCCATCCCTTACCTACAGACAGCTG 1140  
Db 1081 GCGCTGTGCGCCACCAAGATGAGCCGCTGAAGCCATCCCTTACCTACAGACAGCTG 1140  
OY 1141 GAGCTCTAAAGCATTAAGCTGATGAGCTTACCCACAAGTTACCCGAGTCTGTATC 1200  
Db 1141 GAGCTCTAAAGCATTAAGCTGATGAGCTTACCCACAAGTTACCCGAGTCTGTATC 1200



```

|||||
Db 302 TCCCGTGTGCGAGGTGTCCGCGCTGAGCAGGAGCGTGTCCGGAGCTGGCTGTGGCT 361
QY 359 TGGCACAAGAAATGTCAAGCTCTCAACACAGAGAGCTGGCGCTGTGTGGCTACCGGCTCT 418
Db 362 TGGCACAAGAAATGTCAAGCTCTCAACACAGAGAGCTGGCGCTGTGTGGCTACCGGCTCT 421
QY 419 CTGAGCGCCCGGAGAGCTGTGAGCGCCCTCCATTTGACCTGTGTATTTCTTCAACCCAG 478
Db 422 CTGAGCGCCCGGAGAGCTGTGAGCGCCCTCCATTTGACCTGTGTATTTCTTCAACCCAG 481
QY 479 ATGCGTTCTCGGGGCCCCAGGCGCTGACACCGTTTCTTCCCGCATACGAAAGGCCAATG 538
Db 482 ATGCGTTCTCGGGGCCCCAGGCGCTGACACCGTTTCTTCCCGCATACGAAAGGCCAATG 541
QY 539 TGGACCTGCTCCCGAGGGGGGCTCCCGAGGAGACGGGCTGTGGCTGTGGCTGTGGCT 598
Db 542 TGGACCTGCTCCCGAGGGGGGCTCCCGAGGAGACGGGCTGTGGCTGTGGCTGTGGCT 601
QY 599 GCTGGGGGTGTGCGGGGGTCTCTGTGAGCAGGCTGATGTGCGGGCTGTGGAGGCGCTGG 658
Db 602 GCTGGGGGTGTGCGGGGGTCTCTGTGAGCAGGCTGATGTGCGGGCTGTGGAGGCGCTGG 661
QY 659 CTTCGCACTGCTGTGGGCGCTTTGTGTGCGGAGTGTGCGGAGTGTGCTACCCCGCTGG 718
Db 662 CTTCGCACTGCTGTGGGCGCTTTGTGTGCGGAGTGTGCGGAGTGTGCTACCCCGCTGG 721
QY 719 TGAAGTGGCCCGGAGACCCCTGAGACAGGAGACGAGAGGAGCGAGGCGGCTGTGAGG 778
Db 722 TGAAGTGGCCCGGAGACCCCTGAGACAGGAGACGAGAGGAGCGAGGCGGCTGTGAGG 781
QY 779 GCGGGGAGACCCCTGAGACAGGAGACCCCTGAGACAGTGTGTCTCCACAGTGTGAGGCTGTG 838
Db 782 GCGGGGAGACCCCTGAGACAGGAGACCCCTGAGACAGTGTGTCTCCACAGTGTGAGGCTGTG 841
QY 839 GGGGCGTGTGCTGGCGTGTGTGTGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 898
Db 842 GGGGCGTGTGCTGGCGTGTGTGTGTGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 901
QY 899 GCGGCTGGGGGAGACGCTCTCTGCGGGAGCCATTCCTGGGGAGCCGTAAGCGGACATCC 958
Db 902 GCGGCTGGGGGAGACGCTCTCTGCGGGAGCCATTCCTGGGGAGCCGTAAGCGGACATCC 961
QY 959 TCCGCGCGCGGCTTCGCGGCGGAGTGTGAGAGACAGCCTTCTCTTCAAGCAAGAGGCC 1018
Db 962 TCCGCGCGCGGCTTCGCGGCGGAGTGTGAGAGAGCAGCCTTCTCTTCAAGCAAGAGGCC 1021
QY 1019 GCGAGATAGCAGAGAGCTTATCTTCTACAGAGAGTGTGAGAGAGCTGTGAGAGTGTG 1078
Db 1022 GCGAGATAGCAGAGAGCTTATCTTCTACAGAGAGTGTGAGAGAGCTGTGAGAGTGTG 1081
QY 1079 GCGGCGGCTGTGCGGCGGAGCAGGAGCGGAGAGCGGAGCAGCAGCCTTCTTACCTAGAGAGC 1138
Db 1082 GCGGCGGCTGTGCGGCGGAGCAGGAGCGGAGAGCGGAGCAGCAGCCTTCTTACCTAGAGAGC 1141
QY 1139 TGGACGTCCTTAAGCATTAACCTGATGAGCTTACCCACAAGGTTTACCCGAGCTGTGA 1198
Db 1142 TGGACGTCCTTAAGCATTAACCTGATGAGCTTACCCACAAGGTTTACCCGAGCTGTGA 1201
QY 1199 TCCAGCAGCTGGGCTTACCTTCTTCAAGATGAGAGCTGTGAGAGATTCGCAAGTGTGA 1258
Db 1202 TCCAGCAGCTGGGCTTACCTTCTTCAAGATGAGAGCTGTGAGAGATTCGCAAGTGTGA 1261
QY 1259 TGAAGTGGCGGAGAGCGGAGGCTTGTGTAAGTGCAGAAAGGCGAGAAATGAGTGC 1318
Db 1262 TGAAGTGGCGGAGAGCGGAGGCTTGTGTAAGTGCAGAAAGGCGAGAAATGAGTGC 1321
QY 1319 CTGAGGCTCTCTCGGCGGCGCTTCCACAGGCTGTGCGACCTGTGAGCGCTTGTGAGG 1378
Db 1322 CT-----CAAGTGGCGACCGTGTGAGCGCTTGTGAGG 1397
QY 1379 GAAGGGGCGAGCTAGACAAAGACACCTAGACACCTGAGCGGCTTGTACCTGGGTACC 1438
|||||

```

```

Db 1358 GAAGGGGCGAGCTAGACAAAGACACCTTAGACACCTTGACCGCTTCTTACCTGGGTACC 1417
QY 1439 TGTGCTCCCTCAGCGCCGAGAGAGCTGAGCTGCGTCCGCGCCGAGCAGATCGGGGCTCA 1498
Db 1418 TGTGCTCCCTCAGCGCCGAGAGAGCTGAGCTGCGTCCGCGCCGAGCAGATCGGGGCTCA 1477
QY 1499 GGGCCGAGAGCTGTGACACCTGTGACCAAGGAGAGCTGTGAGCTTCTTATCCAAAGGCC 1558
Db 1478 GGGCCGAGAGCTGTGACACCTGTGACCAAGGAGAGCTGTGAGCTTCTTATCCAAAGGCC 1537
QY 1559 GCGTGTCTTCCAGACATGAGAGGCTGTGCAATCTTGTGAGAGATTCAGTCTCTGG 1618
Db 1538 GCGTGTCTTCCAGACATGAGAGGCTGTGCAATCTTGTGAGAGATTCAGTCTCTGG 1597
QY 1619 GTGGGGGCGGAGAGATTTGAGAGGCGCTCAGTACAGAAATGTGAGATGAGTACTGG 1678
Db 1598 GTGGGGGCGGAGAGATTTGAGAGGCGCTCAGTACAGAAATGTGAGATGAGTACTGG 1657
QY 1679 CCAAGTTCATGAAAGCTGTGAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1738
Db 1658 CCAAGTTCATGAAAGCTGTGAGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1717
QY 1739 AACTTGTGGAGCCGAGAGTGTGAGAGGCTGTGAGAGGAGAGAGAGAGAGAGAGAGAG 1798
Db 1718 AACTTGTGGAGCCGAGAGTGTGAGAGGCTGTGAGAGGAGAGAGAGAGAGAGAGAGAG 1777
QY 1799 ACTGATCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1858
Db 1778 ACTGATCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1837
QY 1859 GCATCCCAAGAGGCTTACCTGCTGTAGACCTCAGAGCTGTGAGAGAGAGAGAGAGAGAGAG 1918
Db 1838 GCATCCCAAGAGGCTTACCTGCTGTAGACCTCAGAGCTGTGAGAGAGAGAGAGAGAGAGAG 1897
QY 1919 CTTGCTCTTCTAGACAGCTGTGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1978
Db 1898 CTTGCTCTTCTAGACAGCTGTGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1957
QY 1979 TGGGCTAGAGGCGCCACCTCTGCTGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
Db 1958 TGGGCTAGAGGCGCCACCTCTGCTGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017
QY 2039 GAGACAGAGAGAGGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2098
Db 2018 GAGACAGAGAGAGGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2077
QY 2099 CATGCCCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2138
Db 2078 CATGCCCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2117

```

```

RESULT 4
US-08-426-819A-32
; Sequence 32, Application US/08426819A
; Patent No. 5723318
; GENERAL INFORMATION:
; APPLICANT: Yamaguchi, No. 5723318omi
; APPLICANT: Kojima, Tetsuo
; APPLICANT: Oh-Eda, Masayoshi
; APPLICANT: Hatori, Kunihito
; TITLE OF INVENTION: Genes Coding for Megakaryocyte
; TITLE OF INVENTION: Potentiator
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 197 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..197  
OTHER INFORMATION: /label=fragment  
OTHER INFORMATION: /note="amplified product from pool D using primers 3a51 and  
OTHER INFORMATION: 75a1"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..197  
US-08-426-819A-32

Query Match 9.2%; Score 197; DB 1; Length 197;  
Best Local Similarity 100.0%; Pred. No. 6,7e-34;  
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 AACTCTTGCTGCTCCGCTGCGAGGTGTCGCGCTGAGCAGCAGCGTGTCCGGAGC 346  
DB 1 AACTCTTGCTGCTCCGCTGCGAGGTGTCGCGCTGAGCAGCAGCGTGTCCGGAGC 60  
QY 347 TGGCTGAGGCTTGGCAGAGATGTCATCAAGCTCTCAACGACGACGCTGCGTCTGG 406  
DB 61 TGGCTGAGGCTTGGCAGAGATGTCATCAACGACGACGCTGCGTCTGG 120  
QY 407 CTCACCGGCTCTGAGACCCCGCAGGACCTGAGACGCCCTCCATTGACCTGCTAT 466  
DB 121 CTCACCGGCTCTGAGACCCCGCAGGACCTGAGACGCCCTCCATTGACCTGCTAT 180  
QY 467 TCCTCAACCCAGATGCG 483  
DB 181 TCCTCAACCCAGATGCG 197

RESULT 5  
US-08-426-819A-31  
Sequence 31, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318om1  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihito  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
TITLE OF INVENTION: Potentiator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
US-08-426-819A-31

Query Match 3.5%; Score 75; DB 1; Length 75;  
Best Local Similarity 100.0%; Pred. No. 7.2e-08;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 415 CTCCTGAGCCCCCGAGAGACCTGAGAGCCCTCCATTGACCTGCTATTCCTCAAC 474  
DB 1 CTCCTGAGCCCCCGAGAGACCTGAGAGCCCTCCATTGACCTGCTATTCCTCAAC 60  
QY 475 CCAGATGCTCTCG 489  
DB 61 CCAGATGCTCTCG 75

RESULT 6  
US-08-426-819A-30  
Sequence 30, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318om1  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihito  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
TITLE OF INVENTION: Potentiator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000



Matches 113; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
OY 1632 GGAGGATTGAGGGCGCTCAGTACAGATGTGAGCATGTGGCAGCTTCATGAA 1691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 969 GCAGGAGCAGAGGAGGAGGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1028
OY 1692 GCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1029 GGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1088
OY 1752 CCACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1811
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1089 GGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1148
OY 1812 GCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1149 GGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1192
```

## RESULT 9

```
US-09-130-114-1/c
: Sequence 1, Application US/09130114
: Patent No. 5976807
: GENERAL INFORMATION:
: APPLICANT: Horlick, Robert A.
: APPLICANT: Damej, Bassam B.
: APPLICANT: Robbins, Alan K.
: TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
: TITLE OF INVENTION: From Multiple Transfected Episomes
: FILE REFERENCE: 0867/1D903US1
: CURRENT APPLICATION NUMBER: US/09/130.114
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 5452
: TYPE: DNA
: ORGANISM: VEBNA
US-09-130-114-1
```

Query Match 2.2%; Score 46.4; DB 2; Length 5452;

Best Local Similarity 50.4%; Pred. No. 0.28; Mismatches 111; Indels 0; Gaps 0;

```
OY 1632 GGAGGATTGAGGGCGCTCAGTACAGATGTGAGCATGTGGCAGCTTCATGAA 1691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1836 GCAGGAGCAGAGGAGGAGGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1777
OY 1692 GCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1776 GGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1717
OY 1752 CCACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1811
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 GGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1657
OY 1812 GCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1556 GGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1613
```

## RESULT 10

```
US-08-910-647-1
: Sequence 1, Application US/08910647
: Patent No. 6251433
: GENERAL INFORMATION:
: APPLICANT: Zuckermann et al.
: TITLE OF INVENTION: Compositions and Methods for
: TITLE OF INVENTION: Polynucleotide Delivery
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
```

STREET: 4560 Horton Street

CITY: Emeryville

STATE: California

COUNTRY: U.S.A.

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/910.647

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Fujita, Sharon M.

REGISTRATION NUMBER: 38,459

REFERENCE/DOCKET NUMBER: 1218.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 923-2706

TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 9600 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-910-647-1

Query Match 2.2%; Score 46.4; DB 4; Length 9600;

Best Local Similarity 50.4%; Pred. No. 0.32; Mismatches 113; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
OY 1632 GGAGGATTGAGGGCGCTCAGTACAGATGTGAGCATGTGGCAGCTTCATGAA 1691
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1015 GCAGGAGCAGAGGAGGAGGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1074
OY 1692 GCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1075 GGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1134
OY 1752 CCACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1811
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1135 GGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 1194
OY 1812 GCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1195 GGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 1238
```

## RESULT 11

```
US-07-884-811-15
: Sequence 15, Application US/07884811
: Patent No. 5316921
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
: TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: palin (Genentech)
: CURRENT APPLICATION DATA:
```







Db	2549	GCACGACAGAGAGAGGGCAGCAGCAGGAGGGGCAGAGGGGCAGAGGGGCAGAGGAC	2608
Oy	1692	GCTCGGACGGATGCGGTGCTGCGCTTGCATCTGGCTGAGGTGCAGAAACTTTCGGGACC	1751
Db	2609	GGAGGGGCAGGACAGAGAGAGAGGGGCAGAGGGGCAGAGGGGCAGAGAGCAGAGAGGGCA	2668
Oy	1752	CCACGTGGAGGGCCTGAAAGCGAGGAGCGGCACC GCCCGGTGCGGACTGTGATCTACG	1811
Db	2669	GGACACAGAGAGGGGCAGAGCAGCAGGAGCGGCAGAGCAGAGAGGGGCAGAGGGGCAGAGCA	2728
Oy	1812	GCACCGCAGAGCAGCAGCTGGACACGGTGGGGTGGGGTCTACAGG	1855
Db	2729	GGAGGGGCAGAGAGGGGCAGAGCAGAGAGGGGCAGAGGGGCAGAGGGGCAGG	2772

Search completed: December 7, 2001, 11:20:41  
Job time: 4724 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2001, 09:53:47 ; Search time 2235.52 Seconds  
(Without alignments)  
16761.848 Million cell updates/sec

Title: US-09-215-035-1  
Perfect score: 2138  
Sequence: 1 AGCAATCCGTCGCGGCCGCCCC.....AAAAAAAAAAAAAAAAAAAA 2138

Scoring table: IDENTITY\_NTC  
Gapop 10.0 , Gapext 1.0

Searched: 17159718 seqs, 8763200856 residues

Total number of hits satisfying chosen parameters: 34319436

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:\*

```
1: /cgn2_6/prodata/2/pna/PCRNUS_COMB.seq:*
2: /cgn2_6/prodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/prodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/prodata/2/pna/US08_COMB.seq:*
5: /cgn2_6/prodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/prodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/prodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/prodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/prodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/prodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/prodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/prodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/prodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/prodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/prodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/prodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/prodata/2/pna/US093_COMB.seq:*
18: /cgn2_6/prodata/2/pna/US094_COMB.seq:*
19: /cgn2_6/prodata/2/pna/US095A_COMB.seq:*
20: /cgn2_6/prodata/2/pna/US095B_COMB.seq:*
21: /cgn2_6/prodata/2/pna/US095C_COMB.seq:*
22: /cgn2_6/prodata/2/pna/US095D_COMB.seq:*
23: /cgn2_6/prodata/2/pna/US096A_COMB.seq:*
24: /cgn2_6/prodata/2/pna/US096B_COMB.seq:*
25: /cgn2_6/prodata/2/pna/US096C_COMB.seq:*
26: /cgn2_6/prodata/2/pna/US096D_COMB.seq:*
27: /cgn2_6/prodata/2/pna/US096E_COMB.seq:*
28: /cgn2_6/prodata/2/pna/US097A_COMB.seq:*
29: /cgn2_6/prodata/2/pna/US097B_COMB.seq:*
30: /cgn2_6/prodata/2/pna/US097C_COMB.seq:*
31: /cgn2_6/prodata/2/pna/US098_COMB.seq:*
32: /cgn2_6/prodata/2/pna/US099_COMB.seq:*
33: /cgn2_6/prodata/2/pna/US6000_COMB.seq:*
34: /cgn2_6/prodata/2/pna/US6001_COMB.seq:*
35: /cgn2_6/prodata/2/pna/US6002_COMB.seq:*
36: /cgn2_6/prodata/2/pna/US6003_COMB.seq:*
37: /cgn2_6/prodata/2/pna/US6004_COMB.seq:*
38: /cgn2_6/prodata/2/pna/US6005_COMB.seq:*
39: /cgn2_6/prodata/2/pna/US6006_COMB.seq:*
40: /cgn2_6/prodata/2/pna/US6007_COMB.seq:*
41: /cgn2_6/prodata/2/pna/US6008_COMB.seq:*
42: /cgn2_6/prodata/2/pna/US6009_COMB.seq:*
43: /cgn2_6/prodata/2/pna/US6010_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
44:	2042.6	95.5	2125	65	US-60-324-185-24511
45:	2034.8	95.2	2126	50	US-60-172-660-22854
46:	2034.8	95.2	2126	54	US-60-213-359-6105
47:	1994	93.3	2129	8	US-08-426-819-34
48:	1981.8	92.7	2433	31	US-09-867-701-10869
49:	1976.8	92.5	2289	25	US-09-644-668-9499
50:	1976.8	92.5	2289	25	US-09-652-123-9530
51:	1976.8	92.5	2289	25	US-09-652-127-9551
52:	1976.8	92.5	2289	25	US-09-698-010-11861
53:	1976.8	92.5	2289	27	US-09-699-998-8662
54:	1976.8	92.5	2289	28	US-09-710-280-2725
55:	1976.8	92.5	2289	28	US-09-710-281-4128
56:	1976.8	92.5	2289	28	US-09-710-286-2845
57:	1976.8	92.5	2289	29	US-09-726-171-1716
58:	1976.8	92.5	2289	49	US-60-164-285-821
59:	1962.8	46.4	1108	49	US-60-164-285-821
60:	1962.8	46.4	1108	49	US-60-164-285-821
61:	1962.8	46.4	1108	49	US-60-164-285-821
62:	1962.8	46.4	1108	49	US-60-164-285-821
63:	1962.8	46.4	1108	49	US-60-164-285-821
64:	1962.8	46.4	1108	49	US-60-164-285-821
65:	1962.8	46.4	1108	49	US-60-164-285-821

42	370.4	17.3	441	31	US-09-867-701-2621	Sequence 2621, Ap
43	370.4	17.3	441	53	US-060-207-484-2621	Sequence 2620, Ap
44	365.2	17.1	503	27	US-09-699-998-1770	Sequence 1770, Ap
45	357.8	16.7	442	17	US-09-330-360-151	Sequence 151, App

## ALIGNMENTS

```

RESULT      1
US-60-324-185-24511
: Sequence 24511, Application US/60324185
: GENERAL INFORMATION:
: APPLICANT: Morris, MacDonald
: APPLICANT: Lal, Preeli
: APPLICANT: Diap, Dinh
: TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
: TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
: TITLE OF INVENTION: POLYMORPHISMS IDENTIFIED THEREBY
: FILE REFERENCE: GX-0019-1 P
: CURRENT APPLICATION NUMBER: US/60/324,185
: CURRENT FILING DATE: 2001-09-21
: NUMBER OF SEQ ID NOS: 35862
: SOFTWARE: PERL Program
: SEQ ID NO 24511
: LENGTH: 2125
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: OTHER INFORMATION: Incycle ID NO: 346520.3
US-60-324-185-24511

```

Query Match	95.58;	Score 2042.6;	DB 65;	Length 2125;
Best Local Similarity	98.38;	Pred. No. 0;		
Matches 2086; Conservative	0;	Mismatches 29;	Indels 6;	Gaps 2;

QY	1	AGGAATTCGGGTGGCGGCGACATCCGTCGTGTGACGGGCGGACGAGCAAGCTACCGGT	60
Db	1	aggaattcgggtggcggcgacatccgctcggtgtgacgggcgacgagcaagctaacgggt	60
QY	61	GGACCCAGCGTGGCTCCCTCCCTGGATTACACAGACATAGGCGCTTG--CAACGGCGTCGA	119
Db	61	ggaccacagtggtctccctccctcggtgtctacaagaagccatgacctgccaacggctcga	120
QY	120	CCCCGTTGGTCTCTGTGGGACGCGCC-----TGGACGCTCCTGTTCCTGTCTTCAGC	174
Db	121	ccccgttggtctctgtgagaccgcgcctcgcgacgctctgtctcgtctctcagc	180
QY	175	CTGGATGCGTGCATCCCGCGGAGACCCCTGGCTGAGAAAGACAGGAGAGCTGCGCCCC	234
Db	181	ctcgagatggtgacgcgacctcgagaccctcgtctggaagacagtcagagtcgtgcgc	240
QY	235	CTGGGGGAGTCTGCAGACACCCCCCATACATTTCAGGCTTCGCCCTCCGCAACTCCTT	294
Db	241	ctggaagcgtatctcctgcgcgaaccctaaacttccagctctccctctgcgcacatctct	300
QY	295	GGCTTCCCGTGTGGCGAGGTGTCCGGGCTTGAGCACGGAGGCTGTCCGGAGTACGGCTGTG	354
Db	301	ggcttcccggtgtgtaggggtgtccggcgtgagcagtcgtgtccggaggtcgtgcgtgtg	360
QY	355	GCCATTGGCAGAGAAATGTCMAAGCTCTCAACAGAGCAGCTGGCGCTGTGGGCTACCGG	414
Db	361	gccttggcagagaaagtgtcaagcttcaacagtagagctgtgcgtgtctgcgtccacgg	420
QY	415	CTGTCTGAGCCCCCGGAGAGACTTGAGCGCCCTCCCAATTGACCTGTGCTAATTCCTCAAC	474
Db	421	ctctctgagccccccgagagactgtgacgcctcccatgtgacctgtcgtatctccaaac	480
QY	475	CCAATAGCTTCTGCGGGGCCCCAGAGCTGCAACCCGTTTCTTCTCCGCAACAGGAAGCC	534
Db	481	ccaaatagcttctgtagggccccaagcttgaccggttctctctccgcacaaagaaagcc	540

QY	535	AATGTGACACTGCTCCCGAGAGGGGGCTCCCGAGCCACACGGCTCTCCCTGCGGCTTG	594
Db	541	aatltvgaaccttgcctcccgaaagggggtctcccgaaagcagaagctgtcctcctgcggtcttg	600
QY	595	GCCGTCGAGGGGTGACGGGGGGTCTGTGCTGACGAGAGCGTGAATGCGGGCTCTGGAGGC	654
Db	601	gctctgcgtcgtgtgtgcgggggtctctcgtctgtagaggaagctgatatgtcgggtcctctggagac	660
QY	655	CTGCGTTTCGACACTGCTGGGGCGCTTTGTGAGCCAGTGCGGCCGAAGTGTGCTACCCGG	714
Db	661	ctgagcttgcagacacttgccttgggcgtcttltgtgcgaagtcgagccgaagtgctgtctaacccgg	720
QY	725	CTGGGAGACTCCCGGGAGACCCCTGTGACACAGACACACAGAGAGGACCGAGGGCGGCTTG	774
Db	721	ctgtgtgaacttgcctcggggaaccccttgcacaaagacaaagcaagtcgaagtcgaagtcgtcttg	780
QY	775	CAGGCGGGGGAGACCCCTACAGGACCCCGCTGCACATGGTGTCTCCACGATGGAGCT	834
Db	781	caaggcgagggaacccctctacggccccccgtcgacaatgtgtctgtctccaaagtgtgaagct	840
QY	835	CTGGGGGGCCCTGCGCCGTGAGGGGCGAGCCCATATCCGAGAGATCCGAGGGCGATC	894
Db	841	ctgcgaggcctgcgtccgggtgcgtgtgcgaagccacacacacccgaagatcccgcaaggctc	900
QY	895	GTGGCCGCTGTGGCGGAACCTCTCTCTCGGGACCCATCCTGGCGGCGAGCTGTAAAGGAC	954
Db	901	gtggcgcggtgtgcggcaaacgctctctctcgtggaacccatctctgcgcagctgtgaagcgacc	960
QY	955	ATTCCTCGCGCGCGCTTCGCGGGGAGAGTGAGAAACAGCCCTGTGCTTTCAGAGCAAGAG	1014
Db	961	atctctcgcgcgcgggttccgcgcgggaagtgtgaagaagacagctgtctctcgaagcaagaag	1020
QY	1015	GCCCGCGAGATAGCGAGACCTCATCTTCTACAAACATGTGGACTGTGAAGCCTGCGCTG	1074
Db	1021	gcccgcgagatagaagcggagctctcatctctctaaagaagtgaggagcttgaagcctgcgtg	1080
QY	1075	GATGCGGCGCTGTGGCCACCCAGATGAGACCGCGTAAACGCCATCCCTTACACTAGAG	1134
Db	1081	gattcgcgccttgcgtgcacacacagatgtgaacggtgtgaagcaccatccctctcaactagaag	1140
QY	1135	CAGCTGGACGTCCCTAAGACCTTAAACTGGATGAGACTTACCCACAAGGTTACCCCGAGCT	1194
Db	1141	cagctggaagcgtcccaagaagcaataacgtgaagctctctaaccaagaagttaaccgcgaagct	1200
QY	1195	GTGATCCAGACACTGGGCTACTCTTCTCAGAGTAGAGCCCTGTAGAGACATTGCAATGG	1254
Db	1201	gtgatccgaagcacttgcgtctactctctctctaaagatbagcccttgaagacaattctgcgaatgg	1260
QY	1255	AATGAGAGTCCCTGGAGACCCCTGAAGGCTTCTTGAAGTGTGACAAAGGGACGCAAAATG	1314
Db	1261	aattgtgaagctccctcggagacccctgaaggtcttgcgtgaagtaacaacaagggcagaataag	1320
QY	1315	AGTCTCTAGGCTCTCGGCGGGCCCTGCCAAGTGTGGCCACCCCTGATGCAGCGCTTGTG	1374
Db	1321	agctctcaaggtctctcgtgcggccccctccacaaggtgtgccacctgatctgaacggtcttg	1380
QY	1375	AAGGGAAGGGCCAGCTAGACAAAGACACCTTAGACACCCCTAACCCTTCTTACCTCTGG	1434
Db	1381	aagggaagggtgcagctatgaacaaagacaccccttagaacacctcttcaaccttgg	1440
QY	1435	TACGTCGTCTCCCTACACCCCGAGAGAGTAGTCCGTGCCCCGACGAGCATCTGGGGCG	1494
Db	1441	tacctgtgtctccctcagcccccgaaagatctgaagcttcgtgcctcccaagcagatcttgggtg	1500
QY	1495	GTACAGGCCCCAGACACTGAGACGTGTGACCCAAAGGACAGCTGGAAGTCTCTATCCCAAG	1554
Db	1501	gtcaagcccccaagcacttgcgaacaggtgtgaaccaaggtcagcttgcgtctctatccaag	1560
QY	1555	GCCGCGCTTCTTTCGAGAACATGAACGGGTCCGGAATACTTGTGTAAGTCCAGTCTTC	1614
Db	1561	gcccgccttgccttccagaacatgaagcaggtctcgaataactctgtgtgaagtccaagtcttc	1620

OY 1615 CTGGGGGGGGCCCCACGAGGAGATTGTAAGGCGCTCACTACGACGAANTGACATGAC 1674  
 Db 1621 cggggctggggcccccaaggagattcgaaggcgctcagtcagcaagaatgtagatctgac 1680  
 OY 1675 TTGGCCAGGTTTCATCAATCACTCGGACGATCGGTCGTCGCTTGACTGTGGCTAGAGTG 1734  
 Db 1681 ttggccagcttcatagaactcggaacgtcggaacgtgctgtgtcgcgtttgactctgaggtg 1740  
 OY 1735 CAGAACTTTCGGGAGCCCACTGAGAGGCGCTGAAAGCGGAGAGACGGCACCCGCGGTG 1794  
 Db 1741 cagaactctcggagcccccaagctggaaggcctgaaggcgagagagcgagcacgcgcggtg 1800  
 OY 1795 CGGAGCTGATCTCTACGCGACGCGGACGACGACCTGACAGCTGACAGCTGGGCTAG 1854  
 Db 1801 cggagactgatactcactagcgacgcygagagagcaactctgacacagctcgggctctag 1860  
 OY 1855 GAGCGCATCCCAACGCGCTACCTGGTCTAGACCTCAGCTGACAGAGACCTCTCGGGG 1914  
 Db 1861 ggcggcatccccaagcaagctgactctgtacctcagcaatgcaagaagcgccctctcggg 1920  
 OY 1915 AGCGCTCGCTCTGAGACCTGGACCTGTTCACACCGTCGCGGACCTCTAGCCCTCC 1974  
 Db 1921 agcgctcgctctcctcagagcctggaactgtcttcacagctccggacactgtctcctagcctcc 1980  
 OY 1975 ACCCTGAGCTGAGGCGCCACACTCTCTTGCTGCGCCCAACCTCTGTGGGGATCCCCGCTG 2034  
 Db 1981 accctgactgagggcgcccaactcctctgtcggccccagcctgtcgggataccccgctg 2040  
 OY 2035 GCCAGGACAGGACAGGCGTATCCCGTTCCACCCGAAGAGAACTCGCGCTCAGTAAACG 2094  
 Db 2041 gccaggagcaagcaaggggtgtgtcccgcttcaccccccaagaagaactcgctcagtaaacg 2100  
 OY 2095 GGAACATGCCCCCTGCAGACA 2115  
 Db 2101 ggaacatgccccctgcagaca 2121  
  
 RESULT 2  
 US-60-172-360-22854  
 ; Sequence 22854, Application US/60172360  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Morris, MacDonald  
 ; APPLICANT: Lal, Preeti  
 ; APPLICANT: Diep, Dinh  
 ; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using  
 ; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymorp  
 ; FILE REFERENCE: GX-0007 P  
 ; CURRENT APPLICATION NUMBER: US/60/172.360  
 ; CURRENT FILING DATE: 1999-12-16  
 ; NUMBER OF SEQ ID NOS: 29838  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 22854  
 ; LENGTH: 2126  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; OTHER INFORMATION: Incyte ID No: 346520.3  
 US-60-172-360-22854

Query Match	95.2%	Score 2034.8	DB 50	Length 2126
Best Local Similarity	98.4%	Pred. No. 0		
Matches 2088	Conservative	0	Mismatches 27	Indels 7
			Gaps	3
QY 1 AGGAATTCGGGAGCGGCACTCCGTCGTGTGACGCGCGACAGAGACTACCGT	60			
DB 1 aggaatttcgggagcggcactcccgctgtgtgacgcgcgacagagactaccggt	60			
QY 61 GGACCCAGGTCCTCCTCCTGGGATCTACACAGACCATGCGCTTG-CAACGCGCTGA	119			
DB 61 ggaccacagtgctccctccctccctggatctacacagacacatgcttcgcaacagcgtcga	120			

[illegible]

```
Db 1201 gtgtatccagaccctggtactctctctcaagtgcacctgagacattcgcaagtcg 1260
Qy 1255 AATGTGAGCTCCCTGGAGACCTTGAAGCTTCTTGAAGTCGACAAAGGCGACAATG 1314
Db 1261 aatgtgagctccctggagacccttaagcttcttgtaagtaacaagaagggcacgaatg 1320
Qy 1315 AGTCTCAGAGCTCCTCGGCGCCCTCCACAGAGTGCGCACCTGATGACCGCTTGTG 1374
Db 1321 agtctcagagctctctcgcgccctcccaagtgccacccctgacccgcttcttg 1380
Qy 1375 AAGGGAAGGGGCGAGCTAGACAAAGACACCTAGACACCTGACCGCTT-CTACCTGG 1433
Db 1381 aagggaagggcgacagctagacaagaacaccctagacacccctgacccgcttgtaacctg 1440
Qy 1434 GTACTGTGCTCCCTCAGCCCGGAGAGAGCTAGCTCCCTGCCCCCAGACATCTGGG 1493
Db 1441 gtactgtgctccctcagcccgagagctgagctccgctgccccagacatctgagc 1500
Qy 1494 GGTGAGGCGCCGAGACCTGAGACGTGTGACCCAAAGCAGCTGACGCTCTATCCAA 1553
Db 1501 ggtcagggcccgagacctcgagacagctgtagcccaagcgagctgagcgtctctatccaa 1560
Qy 1554 GGCCCGCTTCTTCCAGAACATGAAACGGGTCCGATATCTTCTGAAGATCCAGTCTCT 1613
Db 1561 ggcccgcttctctcagacaacagaaaggtcgaatactctgtagaagatccagctct 1620
Qy 1614 CCTGGTGGGGGCGCCGCGAGGATTTGAAGCGCTGACAGAGATGAGATGAGATGGA 1673
Db 1621 cctggggtggggcccgcccgagagatctgaaggcgctcagctacagcaatctgaagaatga 1680
Qy 1674 CTGGGCACTTTCAGTAAGCTGCGAGAGGATGCGGTGCTGCTTGAAGTGTGAGAGT 1733
Db 1681 ctgggcaactctctgaagctgagcgagctgagctgctgctgactgtagctgagctgagct 1740
Qy 1734 GCAGAACTTTGTGGAGACCCGACGTGAGAGGCTTGAAGCGGAGAGCGGACCGCGCT 1793
Db 1741 gcagaaactctctgagaccccgctgagagcgctgaaggcgagagcgacgcgcgcgt 1800
Qy 1794 GCGGAGCTGATCTACGAGCGGAGAGAGAGAGCTGAGACGCGTGGGGCTGAGCA 1853
Db 1801 gcggagactgagctcctcagcgagcgagcaagacgacgctgagcgcgagggcgctatac 1860
Qy 1854 GGGGCGCATCCCAACGAGCTACTGTGCTAGACCTAGCGCTGCAAGAACCTCTCCGG 1913
Db 1861 gggcgagactcctcccaagcgctacgctgctctagacctgaagtgagggcgctctcg 1920
Qy 1914 GAGCGCTGCTCTAGAGACTGGAACCTGTTCTACCGCTCTGGCACTGCTCTAGCCTC 1973
Db 1921 gacgcccctgctctcagacctgagacctgtctcaccgctctgacactgctcctagcctc 1980
Qy 1974 CACCTGGGCTGAGAGGCGCCCACTCCTGTGAGGCGCCGAGCTGCTGGGGAGTCCCGGCT 2033
Db 1981 caacctgagctcgagggcgccacccctcctgctgagcccgagcctctcgagggtctcccgct 2040
Qy 2034 GGGCAGAGAGCAGGACGGGTATCCCGCTTCCACCCCAAGAGAACTCGCGTCAAGTAAAC 2093
Db 2041 ggcacagagcagagcgagcggtgctcccgcttcaccccaagagaaactcgcgctcagtaaac 2100
Qy 2094 GGGAACTGCCCTCTGCAGACA 2115
Db 2101 gggaaactgccccctgagaca 2122
```

```
RESULT 3
; Sequence 6105, Application US/60213359
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymor
```

```
; TITLE OF INVENTION: Identified Thereby
; FILE REFERENCE: GX-0015 P
; CURRENT APPLICATION NUMBER: US/60/213,359
; CURRENT FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 7924
; SOFTWARE: PERL Program
; SEQ ID NO 6105
; LENGTH: 2126
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No: 346520.3
US-60-213-359-6105
```

```
Query Match 95.2%; Score 2034.8; DB 54; Length 2126;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 2088; Conservative 0; Mismatches 27; Indels 7; Gaps 3;
```

```
Qy 1 AGAATTCGGTGGGCGGCGGCACTCCGCTGCTGAGCGCGGAGACAGAGCTACCGGT 60
Db 1 aggaatccgggtggcgggccactccgctctgtgagcgcggaagaagctaacggt 60
Qy 61 GGACCCAGGAGTGCCTCCCTCCCTGGGATCTACACAGACCATGCGCTTG-CACGCGCTGA 119
Db 61 ggacccagagtgctccctccctccttgagctacacagacatgaccttgccaagcgctga 120
Qy 120 CCCCTGTGTGCTGTGGGAGCGGCC-----TGGCACCTTCGTCTTCTCTCTACGC 174
Db 121 cccctgtgtggtctgtgtggaaccccgccctcgagcctctctgtctctcctcagc 180
Qy 175 CTCGGAGGAGTACCCGCGAGAGACCTGGCTGAGAGACAGGAGACAGCTTACCGCC 234
Db 181 ctcgagatggtgagcagccctcagagacccctgctgagagaaagcgagagctcgagccc 240
Qy 235 CTGGGGGAGTCTGTACAAACCCCCCATTAATTCACAGCTCTCCCTCGCCAACTCCTT 294
Db 241 ctgagcgagctcgtgccaacccactaacattccacgctctccctcgcacactcctt 300
Qy 295 GGCTTCCCGTGTGCGGAGGTGTCCGGCTGAGACAGAGAGCTGTCCGGAGACTGGCTGTG 354
Db 301 ggtcccgctgtgcgaggtgtctcgccctgagcagcgagcgctgtcggagagctgctgtg 360
Qy 355 GCCTTGGCAGAAAGATGTCAGCTCTCAACAGAGAGCTGAGCTGCTGCTGCTACCGG 414
Db 361 gcttggcagcagaagaatgltcaagctctcaacagagcagcgctgctgtgtctcaagg 420
Qy 415 CTGCTGAGCCCCCGAGAGACTGAGCGCCTCCATTGACTGTGCTATTCCTCAAC 474
Db 421 ctctctgagcccccgagagactggaagcctcccatctgagcctctgtctatctcctaac 480
Qy 475 CCAAGTGGCTTCTGGGGGCGCCAGGCGCTGCAACCGTTTCTTCTCCCGATACAGAGCC 534
Db 481 ccagatggtctctcgggcgcccaagcctgcaacccgtctctctccgatacaagaagcc 540
Qy 535 AATGAGACCTGCTTCGCCGA-GGGGGGCTCCGAGAGAGAGCGGCTGCTGGGCGCT 593
Db 541 aatgtgagactgctcccgagggggggtctcccgagcgaagaagcgctgctcgctgagctct 600
Qy 594 GGCTGTGAGGGGTGTGCGGGGCTCTGCTGAGCGAGGCTGATGTGCGGCTTGGGAGG 653
Db 601 ggctgtgaggggtgtgcggggtctctgtagcgaggtgtagtltgcygggtctcgagggg 660
Qy 654 CCTGGCTTGCACCTGCTGGGCGCTTGTGTGGCGGAGTGGCGGCAAGTGTCTACCCCG 713
Db 661 cctggttgcagctgctctgagcgcttltgtgagcgagctgagcgaaggtcgtacacccg 720
Qy 714 GGTGGTGAAGTGCAGGAGACCTGAGACAGAGACAGAGAGAGAGAGAGGAGGCGGCT 773
Db 721 gctgtgtgagctgcccgggagcccttgagacagagacagagagagcgagcgagctct 780
Qy 774 GCAAGGCGGGGGAGCCCGCTTACGGGCGCCCGCTGACATGCTGTCTCCAGATGAGAGCG 833
```





Query Match	93.3%	Score 1994;	DB 8;	Length 2129;
Best Local Similarity	97.0%;	Pred. No. 0;		
Matches 2075;	Conservative	0;	Mismatches 35;	Indels 30;
				Gaps 3;

QY	5	ATTCCGGTGGCCGGCCCACTCCCGCTCTCTGTACACGCGCGGACAGAGAGCTACCGGTGAC	64
Db	2	AATTCCGGACAGAGCCCACTCCCGCTCTCTGTACAGCGCGGACAGAGAGCTACCGGTGAC	61
QY	65	CCAGAGTGCCTCCCTCCCTGGAGATCTACACAGACCAATGGCGCTTG--CAAGGGCTCGACCC	123
Db	62	CCAGAGTGCCTCCCTCCCTGGAGATCTACACAGACCAATGGCGCTTGCCACAGCGCTCGACCC	121
QY	124	TGTTGGTCTCTGTGGGGACCGGCC---TGCGAGCCTCTGTTCCTCTTACAGCTCG	178
Db	122	TGTTGGGGTCTCTGTGGGAGCCCGGCCCTCGGCGAGCCTCCGTCTCTGCTTACAGCTCG	181
QY	179	GATGGGTGACATCCCGGAGAGACCTGGCTGTGAGAGACAGGAGAGATGTGGCCCCCTGG	238
Db	182	GATGGGTGACACCTCTGAGAGACCTGGCTGTGAGAGACAGGAGAGGTGTGACACCTGG	241
QY	239	GGGAGTCTGTGACAAACCCCATACATTTTCCAGCCTCTCCCTCGCAACTCTTGGCT	298
Db	242	ACGGAGTCTGTGGCCACACCCACCTAATATTCCAGCCTCTCCCTCGCAACTCTTGGCT	301
QY	239	TCCCGTGTGGGAGGTCTCCGGCTGAGCAGGAGCGTGTCCGGAGCTGTGCTGTGGCT	358
Db	302	TCCCGTGTGGGAGGTCTCCGGCTGAGCAGGAGCGTGTCCGGAGCGTGTGCTGTGGCT	361
QY	359	TGGCACAGAAATGTCAACTCTCAACAAGAGAGCGTGGCGCTGTGCTCAACGGCTCT	418
Db	362	TGGCACAGAAATGTCAACTCTCAACAAGAGAGCGTGGCGCTGTGCTCAACGGCTCT	421
QY	419	CTGAGCCCCCGGAGACCTGTGAGACGCCCTCCATTTGAGCCTGCTATTCTCAACCCAG	478
Db	422	CTGAGCCCCCGGAGACCTGTGAGAGCCCTCCATTTGAGCCTGCTATTCTCAACCCAG	481
QY	479	ATGGGTTCTGGGGCCCCAGGCGCTGACACCGTTCTTCTCCGCAATCAGAGGGCAATG	538
Db	482	ATGGGTTCTGGGGCCCCAGGCGCTGACACCGTTCTTCTCCGCAATCAGAGGGCAATG	541
QY	539	TGGACCTGTCCGAGAGGGGGCTCCGAGAGGACAGCGGTGCTCGGCTCTGAGCT	598
Db	542	TGGACCTGTCCGAGAGGGGGCTCCGAGAGGACAGCGGTGCTCGGCTCTGAGCT	601
QY	599	GCTGGGCTGTGCGGGGCTCTCTGCTGACGAGGCTATGTGGGGCTCTGGAGGCTTGG	658
Db	602	GCTGGGCTGTGCGGGGCTCTCTGCTGACGAGGCTATGTGGGGCTCTGGAGGCTTGG	661
QY	659	CTTGCAGACTGTGCGGGCGCTTTGTGCGGAGTGGGCGCAAGAGTGTCAACCGCGCTGG	718
Db	662	CTTGCAGACTGTGCGGGCGCTTTGTGCGGAGTGGGCGCAAGTGTGTCAACCGCGCTGG	721
QY	719	TGAGCTGCCGGGAACCCCTTGAGCAGAGACGACGAGGAGGACCCAGGGCGCTTGCAGG	778
Db	722	TGAGCTGCCGGGAGCCCTTGAGCAGAGACGACGAGGAGGACCCAGGGCGCTTGCAGG	781
QY	779	GGGGGGGAGCCCCCTACGGGCCCCCGCTGCAATGTTCTTCTCAGATGAGACGCTGTGC	838
Db	782	GGGGGGGAGCCCCCTACGGGCCCCCGCTGCAATGTTCTTCTCAGATGAGACGCTGTGC	841
QY	839	GGGGCTCTGTGACCGCTGTGTGGGCAACCATCTCCGACAGATCCCGAGGGCAATCTGGG	898
Db	842	GGGGCTCTGTGACCGCTGTGTGGGCAACCATCTCCGACAGATCCCGAGGGCAATCTGGG	901
QY	899	CCGCGTGGGCGAAGCTCTCTCGGAGACCATCTGTGCGGAGAGCTGAAGGACATCC	958
Db	902	CCGCGTGGGCGAAGCTCTCTCGGAGACCATCTGTGCGGAGAGCTGAAGGACATCC	961
QY	959	TTCGGCCGGGTTTCGGGCGGAGTGGAGAAAGACACCTGTCTCTTACGCAAGAGGCC	1018
Db	962	TTCGGCCGGGTTTCGGGCGGAGTGGAGAAAGACACCTGTCTCTTACGCAAGAGGCC	1021

QY	1019	GGAGATAGACGAGACGCTCATCTTCTACAGAAAGTGGAGCTGGAAAGCTGCGTGGATG	1078
Db	1022	GGAGATAGAGAAAGGCTCATCTTCTACAGAAAGTGGAGCTGGAAAGCTGCGTGGATG	1081
QY	1079	CGGCGCTCTGGCCACCCAGATGGAGCGGCTGTAAGCCATATCCCTTCACACTACAGAGAC	1138
Db	1082	CGGCGCTCTGGCCACCCAGATGGAGCGGCTGTAAGCCATATCCCTTCACACTACAGAGAC	1141
QY	1139	TGGAGCTCTAAAGCATTAACATGGATGAGCTCTACCCACAAAGTTACCCGAGTCTGTGA	1198
Db	1142	TGGAGCTCTAAAGCATTAACATGGATGAGCTCTACCCACAAAGTTACCCGAGTCTGTGA	1201
QY	1199	TCACAGCACTGGGCTACCTCTTCTCAAGATGAGCCCTGAGACATTTGGCAAGTGGAAATG	1258
Db	1202	TCACAGCACTGGGCTACCTCTTCTCAAGATGAGCCCTGAGACATTTGGCAAGTGGAAATG	1261
QY	1259	TGAGCTCCCTGGAGACCCCTGAAGGCTTTGGTAAATGACAAAGGAGCGAAATGAGTGC	1318
Db	1262	TGAGCTCCCTGGAGACCCCTGAAGGCTTTGGTAAATGACAAAGGAGCGAAATGAGTGC	1321
QY	1319	CTCAGAGCTCTCGGCGGCGCCCTCCACAGGTGGCGCCACCCCTGATGACACGCTTTGTGAAG	1378
Db	1322	CTCAGAGCTCTCGGCGGCGCCCTCCACAGGTGGCGCCACCCCTGATGACACGCTTTGTGAAG	1381
QY	1379	GAAAGGCGCCAGCTGAGACAAAGACACCTTAGACACCTGACGCGCTTTACCTTGGATAC	1438
Db	1388	GAAAGGCGCCAGCTGAGACAAAGACACCTTAGACACCTGACGCGCTTTACCTTGGATAC	1441
QY	1439	TGTGCTCCCTAGGCGCGGAGAGCTGAGCGCTGCGCCCGCCAGCAGCAATCTGGGCGGCA	1498
Db	1448	TGTGCTCCCTAGGCGCGGAGAGCTGAGCGCTGCGCCCGCCAGCAGCAATCTGGGCGGCA	1477
QY	1499	GAGCCCAAGAGCTTGAGACGCTGTGACCCAAAGCAGCTGAGACCTCTCTATCCCAAGGCC	1558
Db	1478	GAGCCCAAGAGCTTGAGACGCTGTGACCCAAAGCAGCTGAGACCTCTCTATCCCAAGGCC	1537
QY	1559	GCCTTGCTTTCCAGACATGAAACGGGTCCGAACTACTGTTGTAAGATCCAGTCTTCTCGG	1618
Db	1538	GCCTTGCTTTCCAGACATGAAACGGGTCCGAACTACTGTTGTAAGATCCAGTCTTCTCGG	1597
QY	1619	GTGGGGGCCCCAGGAGGATTTGAAGGCGCTCAGTCAGAGAAATGTAGCATGGACCTTGG	1678
Db	1598	GTGGGGGCCCCAGGAGGATTTGAAGGCGCTCAGTCAGAGAAATGTAGCATGGACCTTGG	1657
QY	1679	CCACGTTTCATGAAGCTCGGAGCGGATGCGGTGCTGCGTTGACTGTGAGCTTAGGTGAGA	1738
Db	1658	CCACGTTTCATGAAGCTCGGAGCGGATGCGGTGCTGCGTTGACTGTGAGCTTAGGTGAGA	1717
QY	1739	AACTTCTGGGACCCACGCTGGAGGGCTCTAAGGCGGAGAGCGCACCCGGTGTGGGG	1798
Db	1718	AACTTCTGGGACCCACGCTGGAGGGCTCTAAGGCGGAGAGCGCACCCGGTGTGGGG	1777
QY	1799	ACTGGATCTTAAGGCAACGGGACGAGGACCTGAGACAGCTGGGGGTGGGGCTTAAGAGGCG	1858
Db	1778	ACTGGATCTTAAGGCAACGGGACGAGGACCTGAGACAGCTGGGGGTGGGGCTTAAGAGGCG	1837
QY	1859	GCATCCCAAGCGGTACTGCTGCTAGACCTCAGGCTGCAAGAGACCTCTCGGGGAGCG	1918
Db	1838	GCATCCCAAGCGGTACTGCTGCTAGACCTCAGGCTGCAAGAGGCGCTCTCGGGGAGCG	1897
QY	1919	CTGCTCTCTAGAGCACTGGAACCTGTTCTACACCGTCTGGACATGCTCTAGCTTCACCC	1978
Db	1898	CTGCTCTCTAGAGCACTGGAACCTGTTCTACACCGTCTGGACATGCTCTAGCTTCACCC	1957
QY	1979	TGGCGCTAGGGCCCCACTCCCTTGCTGGGCCCAAGCCCTGCTGGGATCCCGCTGGCCA	2038
Db	1958	TGGCGCTAGGGCCCCACTCCCTTGCTGGGCCCAAGCCCTGCTGGGATCCCGCTGGCCA	2017
QY	2039	GGAGCAGGCAAGGGTGTATCCCGCTTCACACCCCAAGAGAACTGGCGCTAGTAAACGGGAA	2098
Db	2018	GGAGCAGGCAAGGGTGTATCCCGCTTCACACCCCAAGAGAACTGGCGCTAGTAAACGGGAA	2077
QY	2099	CATGCCCCCTGCAGACAAAAAANAAAAAAAAAAAAAAAAAAAAA 2138	



[illegible]

```

RESULT      6
US-09-644-868-9499
: Sequence 9499, Application US/09644868
: GENERAL INFORMATION:
: APPLICANT: White, David
: APPLICANT: Holtzman, Douglas A.
: APPLICANT: Pan, Yang
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600.1168-001
: CURRENT APPLICATION NUMBER: US/09/644, 868
: CURRENT FILING DATE: 2000-08-28
: PRIOR APPLICATION NUMBER: 60/151, 063
: PRIOR FILING DATE: 1999-08-27
: NUMBER OF SEQ ID NOS: 10075
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9499
: LENGTH: 2289
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-644-868-9499

```

Query Match	Best Local Similarity	Matches 2053; Conservative	92.5%; 97.3%; 0;	Score 1976.8; Pred. No. 0;	DB 25; Matches 27; Indels 30; Gaps 3;	Length 2289;
QY	12	TGGCCGGCCACTCCCGTGTGCTGTATCAGCGCGGAGAGAGGTACCGGNGAGCCACCGT				71
Db	202	tggccggccactcccggtgtgtgtatcagcgcggaagaggtaccggtgagcccaaggt				261
QY	72	GCCTCCCTCCTGGGATACACAGACCATGGCCTTG-CAACGGCTGCACCCCTGTTGGT				130
Db	262	gcctccctccttgatctacacagacacatgaccttgccacagcgtcgagccctgtggg				321
QY	131	CCTGTGGGGAGCGGCC-----TGGAGGCTCCTGTTCCTGCTGCTTCAGGCTGGATGGGT				185
Db	322	gtccctgtggagaccgcccgcctctcgagacctctctctgttccctcttcacgcttggatgggt				381
QY	186	GCATCCCGCGAGAACCTCGCTGGTGGAGAGACAGGAGGTGTGCCCCCTGGGGAGAT				245
Db	382	gcagccctcggagagaccctcgcttggagagacagggcagaagagcttgcgccccttgaacggagt				441
QY	246	CCTGACAAACCCCAATTAACATTTCCAGGCTCTCCCTCGCCAACTCCTTGGGCTTCCCGTG				305

Db	442	ccgycgaaccacccatacatlccacgctccctccgccaactccctgtcccg	501
OY	306	TCGCGAGGCTGTCCGCCCTGAGCAGCGAGCGCTGTCCGGAGAGCTGGCTGTGGCTTGGCACA	365
Db	502	tgccgaaggtgtccgacctgagccacgagcggtgtccgagagctgtgacctgtgcaca	561
OY	366	GAAAGATTTCAGAGCTTTCACACAGCAGAGCTGGCGCTGTGGCTCACCGGCTCTCTGAAGCC	425
Db	562	gaagaaatcgaagctctcaacacagacgacgcgcgtgtccgtctcaacggtctctcgtgacc	621
OY	426	CCCCGAGAGACCTTGAGACCCCTCCCATTTGAGACGCTGCTGATTTCTCAACCCAGATGCTTT	485
Db	622	ccccgaggaactgtgacgctccctccatgtgacccgtcgtcatcttcccaacccaagatgsgt	681
OY	486	CTCGGGGCCCCAGGCTTGACACCGCTTTCTTCTCCGCATACGAAAGCCOATATGSGACT	545
Db	682	ctcgaggcccccaagccgtgacacccgttctctctccgcacatacaagagcccaatgtgtgacct	741
OY	546	GCTCCCGAGGGGGGCTCCCGAGCAGCGCGCTCGCTGGGGGCTTGAGCTGTGGG	605
Db	742	gtctcccgagggggctctcccgagcgacagcggtcgtcgtctcggtctcggtcgtcggtg	801
OY	606	TGTGCGGGGAGTCTGTGCTGAAGCGCTATGTGCGGGCTTGGAGGCGCTTGCTTGCGA	665
Db	802	tgtagcggggtctctgtctgtgacgagcgatgtgtgtgcgggtctctgtgagagccgtgtcgga	861
OY	666	CTGTGCTGGGGGCTTGTGGCGCGAGTCCGCGAAGTGCTGATACCCGGCTGGTGTAGTG	725
Db	862	ctgtgcgtgagcgctctgtgtggcgagctcgtgcgagagtgctgtcaaccccggtcgtgtgagctg	921
OY	726	CCCGGAGCCCTTGAGCAGGACGACGAGCGACGACGAGGCGGCTGTGACAGGCGGGGG	785
Db	922	cccgagacccctgtgacagagacagcgagggagcgaggtgcgtctgtcgagggggg	981
OY	786	ACCCCCATAGGCCCCCGTCCGATGCTGTGCTCCACAGATGAGAGGCTGTGCGGGGCT	845
Db	982	aacccccatacgccccccgtccgaagatgtgtgtctctcaagatgagacgtctcgaggggct	1041
OY	846	GCTGCGCGTGTGGGCCAGGCCCATCATCCGACACATCCCGAGGGGCATGTTGGCGCGTG	905
Db	1042	gtctgcgcgtgtgtggcgaccatcatccgacgcatcccgaggtcatctgtgtgcgcgtg	1101
OY	906	GCGGCAAGCTCTCTCTGGGAGCCCATCTGTGGGCGAGCTGAAAGGACCATCTCCGGCC	965
Db	1102	gcggaaaggtctctctctgtggagccatctctgtgcgcgcctgtgaagcgacctctctgcgc	1161
OY	966	GCGGTTCCGGCGGGAATGTGAGAAGACAGCTCTCTTAGCAGAGAGGCCCGCGAGAT	1025
Db	1162	gcggtctcgggggaaagtgtgagaagacagctgtctctcaagcaagaaggtcccgagat	1221
OY	1026	AGACGAGAGCTCATCTTCTTACAAAGATGTGAGCTGTGAAGCTCTCGTGTGCGGCCCT	1085
Db	1222	agacagagagctctctctctcaagaagatgtgtgtgtcgtgaagctctgtgtgtgcgcct	1281
OY	1086	GCTGGCCACCCAGATGTGAGCCGCGGGAAGCGCATCCCTTACCTTAGCAGAGCGCTGAGAGT	1145
Db	1282	gtctgtccaccccaagtgtgaacccgctgtgaagcgcataccctctcaactacagcgcgctgtgaagt	1341
OY	1146	CTTAAGACATTAACGTGATGAGCTTACCCACAGGTTTACCCCGAGTGTGTGATTCAGCA	1205
Db	1342	ctctaaagataaactgtgagtgtctctaccacaagaagtcttacccgaggtctgtgtgtccagca	1401
OY	1206	CTTGGGCTACTTCTCTCAAGATGTAGCCCTTAGAGACATTTGCCAAGTGTGATGTGAGCTC	1265
Db	1402	ctgtgtgtactctctctctcaagaatgcccctgtgagcatctgtcaagtgtgaatgtgtgaagtc	1461
OY	1266	CTTGAAGACCTTGAAGCTTGTGCTTAAGTGTGACAAAGGGACGAAATGATCTCTCAGGC	1325
Db	1462	ctgtgtgagacctgtgaagcttctgtctgtgaagtcaacaaggtgcggaatgtgaatctctct	1516
OY	1326	TCTCTGGGGGCCCTCTCCACAGGTGCGCACCTGATTCAGCGCTTTGTGAAGGAGAGGGG	1385
Db	1517	-----caagtgtgcacccctgtatctgcagccctctgtgtgaagggaaaggg	1557

QY	1286	CCACCTTGACAAAGAACACCTTAGACACCTTAGACCGCTTTCACCTGGGTACCTGTGTC	1445
Db	1558	CCAGCTAGAAAAAGAACACCCCTAGAACCCCTTAGACCGCTTTCACCTGGGTACCTGTGTC	1617
QY	1446	CCCTAGCCCCGAGAGACTGAGCTCCGTGCTGGCCCCAGACAGCATCTTGGCGGTACAGGCCCA	1505
Db	1618	CCCTAGCCCCGAGAGACTGAGCTCCGTGCTGGCCCCAGACAGCATCTTGGCGGTACAGGCCCA	1677
QY	1506	GGACCTTGACACGTGTGACCCAAAGGACAGTGGACGTCTCTATTCCAAAGGCCCGCTTGC	1565
Db	1678	GGACCTTGACACGTGTGACCCAAAGGAGAGGTGGACGTCTCTATTCCAAAGGCCCGCTTGC	1737
QY	1566	TTTTCCAAACATCATAGCGGGGCCCAATACCTCGTGAAGAATCCAGTCCCTTCGGGGGGGGC	1625
Db	1738	TTTCGAGAACATAGCGGGGCCCAATACCTCGTGAAGAATCCAGTCCCTTCGGGGGGGGC	1797
QY	1626	CCCCACGGAGATTTTGAAGCGCTCAGTCAGCAGAGATGTGACATGATGACTTGGCCAGCTT	1685
Db	1798	CCCCAGGAGATTTTGAAGCGCTCAGTCAGCAGAGATGTGACATGATGACTTGGCCAGCTT	1857
QY	1686	CATGACACTCGGACGGATGCGGTGCTGCCGTTTGACTGTGGCTGAGGTGCGAAGAACTTCT	1745
Db	1858	CATGAACTCTGCGACGAGATGCGGTGCTGCCGTTTGACTGTGGCTGAGGTGCGAAGAACTTCT	1917
QY	1746	GGGACCCCAAGTGAAGGCGCTGAAAGCGGAGAGACGGGACCGCCGGTGGCGGACTGGAT	1805
Db	1918	GGGAGCCCAAGTGAAGGCGCTGAAAGCGGAGAGACGGGACCGCCGGTGGCGGACTGGAT	1977
QY	1806	CTTACGGCAGCGGACGAGACAGCTTGACACGCTGGGGCTGGGGCTTACAGGGCGGCATCC	1865
Db	1978	CTCTAGCAGCGGACGAGACAGCTTGACACGCTGGGGCTGGGGCTTACAGGGCGGCATCC	2037
QY	1866	CAACGGCTACTGTGCTCTAGACCTTCAGCTGCAAGAGAACCTTCCTGGGGAGGCCCTGCT	1925
Db	2038	CAAGGCTACTGTGCTCTAGACCTTCAGCTGCAAGAGAACCTTCCTGGGGAGGCCCTGCT	2097
QY	1926	CCTAGACACTGAGACCTGTCTTCTACCCCTCGTGGCAGCTGCTCTAGACCTTCACACCTTGGCCTG	1985
Db	2098	CCTAGGACCTGAGACCTGTCTTCTACCCCTCGTGGCAGCTGCTCTAGACCTTCACACCTTGGCCTG	2157
QY	1986	AGGGCCCAACTCCCTTGTGCTGGGCCACAGCCCTGCTGGGGATCCCGCGTGGCCAGAGACAG	2045
Db	2218	GGAGCCCAACTCCCTTGTGCTGGGCCACAGCCCTGCTGGGGATCCCGCGTGGCCAGAGACAG	2277
QY	2106	CCTCAGACACA 2115	
Db	2278	CCTCAGACACA 2287	
RESULT 7			
US-09-652-123-9530			
: Sequence 9530, Application US/09652123			
: GENERAL INFORMATION:			
: APPLICANT: Holtzman, Douglas A.			
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
: FILE REFERENCE: 1600.1186-001			
: CURRENT APPLICATION NUMBER: US/09/652,123			
: PRIOR FILING DATE: 2000-08-30			
: PRIOR APPLICATION NUMBER: 60/151,135			
: NUMBER OF SEQ ID NOS: 9796			
: SOFTWARE: FastSeq for Windows Version 4.0			
: SEQ ID NO 9530			
: LENGTH: 2289			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
US-09-652-123-9530			

Query Match Similarity		92.5%	Score 1976.8:	DB 25:	Length 2289:
Best Local Similarity		97.3%	Pred. No. 0:		
Matches 2053:		Conservative	0:	Mismatches 27:	Indels 30: Gaps 3:
QY	12	TGGCCGGCCACTCCCGTGTGCTGTGTAGCGCGGAGAGAGAGCTACCGGTGAGCCAGCGT	71		
DB	202	tgccgcgcacatccctctgtctgtcgtgacgagcagagagctacccggtgagccaccgct	261		
QY	72	GGCTCCCTTCCTGGGATCTACACAGACATGGCCCTTG--CAAGGCTCGACCCCTGTGGT	130		
DB	262	gctccctccctctgagatcatcacagacacatgaccttgccaagcgctcgaccctgtctgg	321		
QY	131	CCTGTGGGGAGACGCCC-----TGCGAGGCTCCTGTTCCTGCTTCCTTCAGACTTGATGGCT	185		
DB	322	gtccctgttgagcccccgccctctgagacctctctgtctccctctacgctctgagatggct	381		
QY	186	GCATCCCGGAGAGACCTGTGGCTGTGAGACAGAGAGAGAGTGTGCCCCCTTGGGGAGT	245		
DB	382	gagacccctcgtagagaccctctgctgagagagacagagagagctctgcgccccctgagagagt	441		
QY	246	CCTGACACACCCCCCAATTAATTTCCAGGCTTCGCCCTGCGCAACTCCCTTGGGTTCCCGTG	305		
DB	442	ccctggcccaaccacacaaatcttcagacctctccctctcgccacatctctgtcttccgctg	501		
QY	306	TGCGGAGGTGTCTCGGCTCTGAGCAGCAGGAGCGTGTCCGGAGCTGAGCTGTGGCTTGAGCC	365		
DB	502	tgcgagaggtctctcgagctcgtgacagagaggtgtctcgggagctgtgctgtgacctgcaca	561		
QY	366	GAAGAATGTCAAGCTCTCAACAGACAGCTGTGGCTGTCTGAGCTTCACCCGCTCTGTGAGGC	425		
DB	562	gaaagatgtccaaagcctctcaacagagacagctgtgctgtctgtctcaacggtctcttgaagcc	621		
QY	426	CCCGAGAGCACTGAGACGCGCTCCACTTGTGAGCTGTGCTGCTATTCTTCACCCAGATGCTTT	485		
DB	622	cccgagagacctgagagcctctccatctgacctgtctgtctatctctcaaccagatgtcgct	681		
QY	486	CTGCGGAGCCCAAGGCGCTGTGACCCGTTTCTTCTTCGCCATTCAGAGAGGCCAATGTGACCT	545		
DB	682	ctcggggccccagagcctgtgacacccgtctctctcccgcatcaacgaaggtccaatgtgagacct	741		
QY	546	GCTGCCGAGGGGGGCTCCCGAGCAGACAGAGGGGTGCTGCGGGGCTCTGGGCTCTGGGCTGGG	605		
DB	742	gtctcccgagaggggtctcccgagagacagaggggtctgtctgctgctgctgtgctgtgggg	801		
QY	606	TGTGCGGGGCTCTGTGCTAGGAGCGCTGATGTGCGGCTCTGGAGAGCGCTGAGCTTGCGA	665		
DB	802	tgtgctgggggtctctctgtgtcgtgacgtgagctgtgtgtgctgtctggagggcctgtgtgca	861		
QY	666	CCTGCTGTGGGCGCTTGTGTGTGGCCGAGTCCGCGCAGATGTGTCTACCCGGGCTGTGAGCTG	725		
DB	862	ctgtctgtgggtctgtgtgtgctgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	921		
QY	726	CCCGGAGACCTTGAGACAGAGAGCAGAGAGGAGCAGCAGAGGCGGCTCTCAGAGGCGGGGG	785		
DB	922	cccgagagacccctcgagacagagacagcagagaggtgagcaggggtgtctgtcagagggcggggg	981		
QY	786	ACCCCTTAAGGCCCCCGCTGTGATGTTGTCTGTCTTCACGATGAGACGCTGTGCGGGGCT	845		
DB	982	acccccctacgtgcccccgctgacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1041		
QY	846	GCTGTCCGTGTGTGGGCGACACCCATCATTCGCGACATCCCGGAGAGGAGATGTGCGCGCTG	905		
DB	1042	gtctgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1101		
QY	906	GGGGAACGCTCTCTCGGGACCAATCTATGGGCGGACGCTTGAGAGGAGCCATCTCCGGGCC	965		
DB	1102	gtgtgcaagctctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1161		
QY	966	GGGTTTCGGCGGAGGTGTGAGACAGACACGCTGTCTTTAGGCAAGAGGCCCCGCGAAT	1025		
DB	1162	gtgtgtcttcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1221		

```
QY 1026 AGACGAGACCTCATCTTCTACAGAAAGTGGAGCTGGACGCTGCGTGAGCGCCCT 1085
    |||||||
Db 1222 agagagagagctctctctctcaagaagtggagctcgaagcccgctgctgagccct 1281
QY 1086 GCTGGCCACCCAGATGAGACCGCGTGAACGCCATCCCTTCACTACAGACAGCTGAGCGT 1145
    |||||||
Db 1282 gctggccacccagatgagccgctgaaagccatccctctcaacgagcagctggagct 1341
QY 1146 CCTAAGGATTAAGTATGATGCTCTACCCCAAGGTTACCCCGAGTGTGATTCACAGA 1205
    |||||||
Db 1342 cctaagataaacttgatgagctctcaacaaagcttaacccgagctctgattccagca 1401
QY 1206 CCTGGGCTACCTTCTCTCAAGATGAGCCCTGAGACATTTGCAAGTGGATGTGACTTC 1265
    |||||||
Db 1402 cctgggctactctctctcaagaatgagccctgaggaactctgcaagtggaatgagcgtc 1461
QY 1286 CCTGGAGACCTGAAAGCTTTGCTTGAAGTGCACAAAGGACGAAATGATCTCAAGC 1325
    |||||||
Db 1462 cctggagacccctgagagctctgcttgaaatgaaacaaagggcagaaatgagctctct 1516
QY 1326 TCTCTGGGGGCGCCCTCCACAGGTGGCCACCTGATGACCGGCTTGTGAAGGGAGGG 1385
    |||||||
Db 1517 -----caggtggccacccctgactgacccgctctgaaaggaaagg 1557
QY 1386 CCAGCTAGACAAAGACACCCCTAGACACCCCTGACCGCTTCTACCTGGGTACTGTGCTC 1445
    |||||||
Db 1558 ccagctagacaagaacacccctagacacccctgacccgctctctacccctggtactcgtgtc 1617
QY 1446 CCTAGACCCCGAGAGAGCTGAGACTCCGTCGCCCCAGCAGCATCTGGCGGTGACGCCCA 1505
    |||||||
Db 1618 cctaagccccgagagagctgagctctgccccagcagcactctgggggtgcaaggcccca 1677
QY 1506 GGACTGTGACAGCTGTGACCCCAAGGACAGCTGAGCTCCCTATCCCAAGGCGCGCTTGC 1565
    |||||||
Db 1678 ggaactggaacacggtgagcccaagacgctgagctctctacccaaagcccgctctgc 1737
QY 1566 TTTCAGAAATGAACGGGCTCCGAATCTCTGTAAGATCCAGTCTTCTGCTGGGGGG 1625
    |||||||
Db 1738 ttctcagaacatgaaaggtctcgaactctgtaagatccagctctctcgtggggg 1797
QY 1626 CCCACGAGAGATTTGAAAGGGGCTCAGTCAACAAATGAGCATGAGCTGGGACGCTT 1685
    |||||||
Db 1798 cccacgagagatcttgaaagcgctcagtcagcaaaatgagcaatgagctctgccaagt 1857
QY 1686 CATGAAGCTGGGAGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1745
    |||||||
Db 1858 catgaagctggagacgagatgagctgctgctgctgctgctgctgctgctgctgctgct 1917
QY 1746 GGGACCCACCTGAGAGGGCTGAAAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGT 1805
    |||||||
Db 1918 gggagcccaagctgagagggctgaaagggcagagggcagagggcagagggcagaggt 1977
QY 1806 CCTACGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 1865
    |||||||
Db 1978 cctacggagagagagagagagagagagagagagagagagagagagagagagagagag 2037
QY 1866 CAAGGCTACCTGCTCTAGACTCAGCGTCAAGAGAGCCCTCTCGGAGAGCCCTGCT 1925
    |||||||
Db 2038 caaagctactgctctcagagccctcagatgcaagagggcctctcggggagagccctgct 2097
QY 1926 CCTAAGACTGAGACTGTTCTACCGTCTGGAAGTGTGCTTACGCTTCAACCTGGGCTG 1985
    |||||||
Db 2098 cctagagactgactgctctcagagccctcagatgcaagagggcctctcggggagagccctgct 2157
QY 1986 AGGAGCCACCTCTTGTGCTGGCCCGAGCCCTGCTGGGATATCCCGCTGGCCAGAGAG 2045
    |||||||
Db 2158 agggagcccaactctctgctggcccaagccctgctggggatcccgccctggcagagagag 2217
QY 2046 GCAGGGGATCCCGTTCAGCCCAAGAGAACTCGGCTCAGTAAAGGGAGAAATGAGCC 2105
    |||||||
Db 2218 gcaaggggtgctcccgctcagcccaagagaaactcgcctcagtaaaagagaaactgccc 2277
```

```
QY 2106 CCTGCAGACA 2115
    |||||||
Db 2278 cctgcagaca 2287

RESULT 8
US-09-652-127-9551
; Sequence 9551, Application US/09652127
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600.1183-001
; CURRENT APPLICATION NUMBER: US/09/652.127
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/151.134
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 10475
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9551
; LENGTH: 2289
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-652-127-9551

Query Match 92.5%; Score 1976.8; DB 25; Length 2289;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2051; Conservative 0; Mismatches 27; Indels 30; Gaps 3;

QY 12 TGCGGGGCACTCCGCTGCTGTGAGCGCGGAGAGAGACTACCGGTGAGCCAGCGT 71
    |||||||
Db 202 tggcgggcaactcccgctgctgtgagcgcgcgagagagactacccgtgagccacagct 261
QY 72 GCGTCCCTCCCTGGGATCTACAGAGACCATGCGCTTG-CAACGGCTGAGCCCGTGTGT 130
    |||||||
Db 262 gctccctccctgggattacagagaccaaagcttgcaagagctcgaaccccgcttggg 321
QY 131 CCGTGGGAGACCGCC-----TGCGAGCTCTGCTGCTTCTTCAAGCCTGAGATGGT 185
    || || ||
Db 322 gtctgtggagcccccgcctccgcgcgcgcctctctctctcagcctcggatgggt 381
QY 186 GCATCCCGCCAGAGACCTGCTGTGAGAGACAGGAGAGAGTCTCCCCCTGGGGGAGT 245
    || || ||
Db 382 gcaagccctcgaagaccccgctgagagagagagagagagagagagagagagagag 441
QY 246 CCTGACAAACCCCATTAATTCAGCTTCCAGCTTCCCTCCGCAACCTGCTGCTCCCGT 305
    || || || || ||
Db 442 cctggcaaccccaactaaacttccagctctccctccgcaactctctgctcccggtg 501
QY 306 TGGGAGGTGTCGGGCTGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 365
    |||||||
Db 502 tgggaggtgtccggcctgagacagagagcggtgtccggagagctgtgtgtgtgtgtgt 561
QY 366 GAAGAATGTCAAGCTTCAACAGAGACACTGCTGTGTGCTACCGGCTCTGTAGACC 425
    |||||||
Db 562 gaagaatgtcaagcttcaacagagagcgctgtgtgtgtgtgtgtgtgtgtgtgtgt 621
QY 426 CCCGAGGAGACTGGAGCGGCTCCCATGAGACTGCTGTGTATTCATCAACCATGAGTT 485
    |||||||
Db 622 ccccgagagactggagcgccctcccatggagactgtgtgtgtgtgtgtgtgtgtgtgt 681
QY 486 CTCGGGGCCCGAGGCTGACCCGCTTCTTCTCCGATCAAGAGGCAATGAGACT 545
    |||||||
Db 682 ctccggggcccgagcctggaacccgtctctctctccgcatcagagagcgaatgtggact 741
QY 546 GCTCCGAGGGGGGCTCCGAGCAGAGCGGCTGCTGCTGCGGCTGCTGCGGCTGCTGGG 605
    |||||||
Db 742 gctcccgagggggtctcccgagcgagcgagcgctgtgtgtgtgtgtgtgtgtgtgtgt 801
QY 606 TGTGCGGGGCTCTCTGCTGAGCGAGGCTGATGTGCGGGCTTGGGAGGCTTGCTGGA 665
    |||||||
Db 802 tgttcgggggtctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 861
```









US-09-699-998-8662

Query Match 92.58; Score 1976.8; DB 27; Length 2289;  
Best Local Similarity 97.38; Pred. No. 0;  
Matches 2053; Conservative 0; Mismatches 27; Indels 30; Gaps 3;

```
Oy 12 TGGCGGCGCACTCCCGTCTGTGAGCGCGGACAGAGAGCTACCGGTGAGCCAGCT 71
Db 202 Tggcgggccacatcccgctcgtctgacgcgagacagagctacggtgagccacggt 261
Oy 72 GCCCTCCCTCCGTGGATCTACACAGACCATGGCCTTG-CAAGGGCTGAGCCCTGTTGT 130
Db 262 gccctccctccctggatctacacagacacatgaccttgcaacgctcagccctgttgg 321
Oy 131 CCTGTGGGAGACCCG-----TGGCAGCTCTCTTCCCTGCTTTCAGACCTGGATGGGT 185
Db 322 gctcgtgaggaccccgccctcgtcagctcctcgtctcctctcctcagcctcgatggt 381
Oy 186 GCATCCCGCGAGACCTGGCTGGAGAGACAGGAGAGTCTCCCGCTGGGGGGAGT 245
Db 382 gcagccctcctgagacccctgctgagagacagagcagagcgtgcccctgagcaggt 441
Oy 246 CCTGACAAACCCCATTAATTTCCAGCCTCTCCCTGCGCAACTCTGTGCTTCCGCTG 305
Db 442 cctgacacacacactaacattccagcctcctccctcgccaactccttgctccggtg 501
Oy 306 TGGCGAGGTGTCCGCGCTGAGCAGAGAGCGCTGTCCGGAGTGGCTGGCTTGGCACA 365
Db 502 Tggcgagggtgctcggcctgagcagagcgtgtccggaggctggtggtccttgagaca 561
Oy 366 GAAGAATGTCAAGCTTCACAGAGACAGCTGCTGTGCTGCTCAGCGGCTCTCTGAGCC 425
Db 562 gaagaaagtcaagctctcaacagagacagcgtgctgctgctcaacggtcctctgagcc 621
Oy 426 CCCCAGAGACTGTGAGCGCCTCCCATTTGACCTGCTCTATTCCTCAACCCAGATGGTT 485
Db 622 ccccgagacactgagcgccctcccatctgagcctgctatctcctcaacccagatgctgt 681
Oy 486 CTGCGGGCCCCAGGCGCTGACACCCCTTTCTTCTCCCGCATACAGAGCGCATGTGAGCT 545
Db 682 ctcgggcccccaaggcccgacccgcttctctcctcgatcagaaagtcgaatgtgagct 741
Oy 546 GCTTCCCAAGGGGGGCTCCCGAGCAGACAGCGGCTGCTGCTCGGCTGTGAGCTTGGAG 605
Db 742 gctcccgagggggtctcccgagcgacagcgtgctgctcggtcctggtcctctgggg 801
Oy 606 TGTGCGGGGCTCTGTGAGCGAGGCTGATGTGCGGCTCTGGAGGCTGAGCTTGGCA 665
Db 802 Tgtgcgggggtctctgctgagcgaggtgatgtgctgagggctctggagggcctgctgca 861
Oy 666 CTTGCGGGGGCGCTTTGTGCGCGAGTGGCGGAGTGTGCTACCCGGCTGGTGAAGTGTG 725
Db 862 cctgctcgggggtctctgctgagcgaggtcggaagtgtctgtacccggctggtgagctg 921
Oy 726 CCCGAGACCCCTGTGACAGAGACAGCAGAGAGGAGCGGCTGTGAGGGCGGGAG 785
Db 922 ccgaggaccccttgagccaagagacagagagcagggcaggtcctctcgagggcgggggg 981
Oy 786 ACCCCCTTACGGCCCCCGTGCATGTGTCTTCCACGATGAGCGCTTGGGGGGCT 845
Db 982 accccctacggcccccgctgacatgtctgtctcaagatgtagcgctctcgggggcct 1041
Oy 846 GCTTCCGCTGTGGGCGAGCCCATATCCGAGAGATCCCGAGGGCATCTGGCGCGGTG 905
Db 1042 gctgctcggtgctgaggacgacatctcgcagatcccggaaggtcatctggtcgcgctg 1101
Oy 906 GCGCAACGCTCTCTGTGGGAGCCCATCTGTGGGAGAGCTGAAGGACCATCTCGGCGC 965
Db 1102 gcgggaaagctcctctcgtggagccatctctggcgagccttgaagagacatcctctcgcc 1161
Oy 966 GCGGTTCCGCGGAGAGTGAAGACAGCCTGTCTTCAAGCAAGAGGCCCGCGAGAT 1025
Db 1162 gcggttccggtcgaggatgagaaagacagccgtctcctcaagcaagaagcccgagat 1221
Oy 1026 AGAGAGAGCTCATCTTTTACAGAACTGGGAGCTTGGAACTTGGATGGGGCCT 1085
Db 1222 agagagagctcatctctctcaagaagtgtagctggaagcctgctgtgagtgagccct 1281
Oy 1086 GCTGGCCACCCAGATGGAGCGCGTGAACGCGCATCCCTTCCTACCTACGAGCGTGAAGCT 1145
Db 1282 gctggccccagatgagccgctggaacgcaatccctctcaactcaagagcagcgtgaggt 1341
Oy 1146 CCTAAGCATTAAGTGAATGAGCTTACCCACAAGTTACCCCGAGTCTGTGATCCAGCA 1205
Db 1342 cctaaagcataaacttgatgagctctaacccaaggttaaccgcgagcttgatgacagca 1401
Oy 1206 CTTGGGCTACTCTTCTCTCAAGATGAGCCCTTGAGAGACATTGGCAAGTGGATGTGAGCTC 1265
Db 1402 cctggtcactcctctcctcaagaatgagccctggaacatctgcaagtggaatgtagagctc 1461
Oy 1266 CCTGAGACCCCTGAAGGCTTGTGAAAGTGCACAAAGGCGACGAAATGAGTCTCAGGC 1325
Db 1462 cctgagagacccctgagaggtcttgcttgagatcaacaagggcagcgaatgagctc----- 1516
Oy 1326 TCTTGGCGGCCCTCCACAGGTGGCCACCTGATGACCCGTTTGTGAAGGAAGGGG 1385
Db 1517 -----caggtggcacccttgatcgaccgcttctgtgaagggaagg 1557
Oy 1386 CCAGCTAGACAAAGACACCCAGACACCTTGAGCGGCTTCTACCTGGGTACCTGTGCTC 1445
Db 1358 ccagctagaaacaaagacacccatagacacccgtgacccctctcaaccctgggtacccgtgtc 1617
Oy 1446 CCTCAGCCCCGAGAGAGTGAAGCTCGTGTGCCCCAGCAGCATCTGCGGCGGTCAAGCCCCA 1505
Db 1618 cctagaccggagagatgtagctcgtgctgctcccgacagcatctggggcggtcagagcccca 1677
Oy 1506 GGACCTGGACACGTGTGACACCAAGCAGCTGAGCTCTTATCCAAAGCCGCCCTTGC 1565
Db 1678 ggaacctgacacgtgtgacccaagcgagctgagacgtctctctatcccaagcgccgctgtgc 1737
Oy 1566 TTTGCGAAGCATGAACGGGTCCGAATCTTGTGAAGATCCAGTCTTCCCTGGGTGGGGC 1625
Db 1738 tctccagaagaatgaaaggggtcgcgaatctcgtgaaatccagctctccggtctgggggc 1797
Oy 1626 CCCCAGGAGGATTTGAAGGCGCTCAGTCAACAGATGTGAGCATGTGACTTGGCCAGCTT 1685
Db 1798 ccccaagagagatcttgaaaggggtccagtcagcaagaatgtagcatgtgacagctt 1857
Oy 1686 CATGAAGCTCGGAGCGGATGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1745
Db 1858 catgaagctcgagcagatgctggtgtgcgtctgacgtgtgaggtgtgcagaaactct 1917
Oy 1746 GGGACCCACGTGAGGCGCTGAGGCGGAGAGCGGACCGCGGCTGGCGGACTGGAT 1805
Db 1918 gggaccccaagctgaaagggctgaaagggcgaagagcgcgaaccccggtggtgagctggat 1977
Oy 1806 CCTACGGGAGCGGAGAGCAGCAGCTGACACCTGAGGCTGAGGCTTACAGGCGGCTATCC 1865
Db 1978 cctacggcaggtgagagagcagccctggaacgcttggtggtctgagctaaagggcgcatccc 2037
Oy 1866 CAACGCTACTGCTCTAGACCTTACGCTGCAAGACACCTTCTGGGGAGCGCCTGCT 1925
Db 2038 caaaggtcactctgtctcttaagaccttaagcatgtcaagaggtccctctcgtggagacgctcct 2097
Oy 1926 CCTAGGACCTTGAACCTGTTCACCGTCTGAGCACTGCTTCTTACGCTTCCACCTCGGGCTG 1985
Db 2098 cctaggaacctgagacctgtctcaacggtctcggaactgtctctcaagctctcaacccgtgctg 2157
Oy 1986 AGGCGCCACTTCCCTTGTGCTGAGCCAGCCCTGTGCTGGGATCCCGGCTGTGAGAGAG 2045
Db 2158 agggccccacacctcttgcttgccccagccctgctgggagatcccgccgtgagagagag 2217
Oy 2046 GCACGGGTGATCCCGCTTCCACCCCAAGACAGACTGCGCTTGAAGTAAAGGAGCATGCGCC 2105
Db 2218 gcaaggtgtgtcccgcttccaccccaagagaaactgcgtctagtaaaaggaacatagccc 2277
```

```
QY 2106 CCTGCAGACA 2115
      |||||
Db 2278 cctgcagaca 2287

RESULT 11
US-09-710-280-2725
: Sequence 2725, Application US/09710280
: GENERAL INFORMATION:
: APPLICANT: Holtzman, Douglas A.
: APPLICANT: Geating, David P.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: FILE REFERENCE: 1600.2007-001
: CURRENT APPLICATION NUMBER: US/09/710,280
: PRIOR FILING DATE: 2000-11-10
: PRIOR APPLICATION NUMBER: 60/164,537
: NUMBER OF SEQ ID NOS: 2763
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2725
: LENGTH: 2289
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-710-280-2725

Query Match      92.5%; Score 1976.8; DB 28; Length 2289;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2053; Conservative 0; Mismatches 27; Indels 30; Gaps 3;

QY 12 TGGCGGCGCATCCCGTGTGCTGAGCGGAGACAGACTACCGGTGAGCCAGCGT 71
      |||||
Db 202 tggcgcgcaatcccgctgtctgtgagcgcgagacagagactacggtgagccaggt 261

QY 72 GCTTCCTCCTCGGATGTACAGACCATGCGCTTG-CAACGGCTCAGCCCTGTGTT 130
      |||||
Db 262 gcttcctcctcgatgtactacacagacatgctcgaacggtcgaacccctgtg 321

QY 131 CCTGCGGAGACGGCC-----TGGCAGCGTCTGTTCTGCTTTCACCTCGATGAGT 185
      |||||
Db 322 gtccgtggaaccccgccctcgagcctcctgtctcctcctcagcctcgatgagt 381

QY 186 GCATCCCGCGAGCCCTGCGCTGAGAGACAGGACGAGTCTGCCCTGCGGAGAGT 245
      |||||
Db 382 gcatcccgcgagccctcgctgtgagagacagagagagagctgcccctggaagagt 441

QY 246 CCTGACAACCCCATTAACATTTCCAGCTCTCCCTCGCAACTCTTGGCTTCCGTG 305
      |||||
Db 442 cctgccaaccccatlaacattccagcctcctccctcgcaactccttgctccgtg 501

QY 306 TGGGAGAGTTCGGGCTGAGCAGGAGCGTCCGGAGCTGGTGTGGCTTGGACACA 365
      |||||
Db 502 tgggagagtgtccggtcgagcagagagctgtccggtgagctgtggtccttgaca 561

QY 366 GAAGAATGTCAAGTCTTCAACAGACAGCTGCGTGTGCTCACCAGCTCTGAGACC 425
      |||||
Db 562 gaagaatgtcaagcttcaacagacagagcgtcgtcgtcgtcagcgtcctcctgagcc 621

QY 426 CCGCGAGAGCTGAGCGCCCTCCATTGGACCTGCTGCTATTTCCTCAACCCAGATGCGTT 485
      |||||
Db 622 cccgagagactgtgagcgtccctcccatgtgacctgtctatctcccaaccagatgagct 681

QY 486 CTCGGGGGCCAGGCGCTGACCCGTTTCTTCCCGCATCAGGAAGGCAATGTGAGCT 545
      |||||
Db 682 ctccggggccagagcgtcgtcacccgttctcctccgcatcaggaagcgaatgtgacct 741

QY 546 GCTCCGCGGAGGCGCTCCGAGCGAGCAGGCGTGTGCTCGCTCGGCTGTGGGCG 605
      |||||
Db 742 gctccgcgagggcgctcccgagcgagcagagcgctcgtcgtcgtcgtcgtcgtcgtcgt 801

QY 606 TGTGCGGAGGTCTGTGAGCAGGAGCTGATGTGCGGCGCTCTGGAGGCGCTGCTTGC 665
      |||||
```

```
Db 802 tgtgaggggtctctgtcgtgagcgaagctgatlgtcgaggtcctcgtgagggcctgcttgcga 861
      |||||
QY 666 CCGCTGCGGCGCGCTTTGTGCGCGAGACTCGCGGAGAGTGTGCTACCCCGCTGTGAGAGT 725
      |||||
Db 862 cctgccttgagcgtcttgtggtccgaatcgtccgaagtgtctgtaccccggtcgtgtgagctg 921

QY 726 CCGGAGACCCCTGAGACACAGCAGCAGAGCAGCAGGCGGCTGTGAGGCGGAGG 785
      |||||
Db 922 ccggagacctctgagaccagagccagagagcagcgagcggtcctcgtcagggcgaggg 981

QY 786 ACCCCCTACGAGCCCCCGGTGACATGCTGTCTCAGATGGAAGCTGTGCGGGGCT 845
      |||||
Db 982 accccctacgcccccgctgacatgtctgtctccacagatgagcgtcgtcgaggcct 1041

QY 846 GCTGCGCGTGTGCGGCGCGCCATCATTCGAGAGATCCCGAGGCGCATCGGCGGCTG 905
      |||||
Db 1042 gctgcccgtgtggtccagccatcgtccagagatcccgagggcatctgtggtccgtg 1101

QY 906 GCGCAACGCTCTCTCTCGGAGCCCATCTGCGCGAGCCTGAGAGCATCTCTCGGCG 965
      |||||
Db 1102 gcggaacagctcctctcgtgagcccatccttggcgagcgtcgtgaagacatcctccgtgc 1161

QY 966 GCGTTCGCGGCGGAGAGTGAAGACAGAGCTGTCTTTCAGGCAAGAAGGCCCGAGAT 1035
      |||||
Db 1162 gcggtccggtcggtgaggtggaagagacagcctgtcctcaggtcgaagagcgccgtgagat 1221

QY 1026 AGACGAGAGCTCATCTTCTCTCAAGAACTGGAGACTGGAACCTGCGCTGTGCGGCGCT 1085
      |||||
Db 1222 agacgagagctcatctctcacaagaatgtagatgtgagcgtcggtgtgtgagccct 1281

QY 1086 GCTGCGCACCCAGATGAGACCCGCTGAACGAGCCATCCCTTCACTACGAGCAGTGTGAGAGT 1145
      |||||
Db 1282 gctggtcccccagatgtgagccgtgtgaacgcatactccctcactcagagcagcgtgagagt 1341

QY 1146 CCTAAAGCATTAACCTGATGAGCTTACCCACAAGATTACCCCGAGTGTGTGATTCACA 1205
      |||||
Db 1342 cctaagaataaactgtatgtgactctacccaagaagttaaccgagttgtgtgactcagca 1401

QY 1206 CCGGCGCTACCTCTTCCCAAGATGAGCGCGTGAAGACATTTGGCAAGTGAAGTGAAGT 1265
      |||||
Db 1402 cctgggtcaacctctcctcaagaatgagccctgtgagaaactctgcaagtgagatgtgagctt 1461

QY 1266 CCGTGAAGCCCTGAAGGCTTTGCTTGAAGTGAAGACAAAGGCGACGAATAGATCTCTCA 1325
      |||||
Db 1462 cctggtgagccctgtgaaggtctgtgtgagttcaacaagaaggaacgaatgtgctcct----- 1516

QY 1326 TCTTCGGCGGCGCTCCACAGGTGCGACCTGATCGACCGCTTTTGTGAAGGGAAGGG 1385
      |||||
Db 1517 -----caggtgtgccaacctgtatcgaccgtcttgtgaagggaagggtg 1557

QY 1386 CCAGCTAGACAAGACACCTAGACACCCCTTACCGCTGGGTACCTGTGCTC 1445
      |||||
Db 1558 ccaagtaacaagaacacacacagacacccgtgacgctctcactccctgtgtaacctgtgtctc 1617

QY 1446 CCTCAGCCCCGAGAGAGCTGAGCTCGTGTGCCCCCAGCAGCATCTGTGGGGGTCA 1505
      |||||
Db 1618 cctcagccccgagagagctgtgagctcgtgtcccccaagcagatctcgtgggtgtcagggcccca 1677

QY 1506 GGACCTGAGACAGTGTGACCCCAAGGACAGTGTGCTCTATCCCAAGGCCCGCTTGG 1555
      |||||
Db 1678 ggaactgtgacaagtgtgagcccaagagcgtgtgagcgtcctctatcccaagcgccgttgc 1737

QY 1566 TTTTCAGAACTGAAGCGGCTCCGAATACTTGTGTGAAGATCCAGTCTTCTGTGGGTGG 1625
      |||||
Db 1738 ttccagaacaatgaacggtgtccgaatacttctgtgaagatccagtcctcctcgtgtgtggtg 1797

QY 1626 CCGCAGGAGATTGTAAGGCGCTCAGTACAGATATGAGATGTGAGCTGTGGCACACTT 1685
      |||||
Db 1798 ccccaagagagatlttgaaagcgtcagtcagcaagaatgtgtgagatgtgacatgtgccaagtt 1857

QY 1686 CATGAAGCTGCGAGCGATGCGGTGCTGCTGCTTGAATGTGTGAGTGAAGTGAAGTCTTCT 1745
      |||||
```



```
QY 1326 TCTGTGCGGCCCTCCACAGATGAGCCCTGATGACCGCTTTGTGAAGAGAGGGG 1385
Db 1517 -----cagtggtccacccttgatcgaccgcttggagaggaagggg 1557
QY 1386 CCAGCTAGACAAAGACACCTTAGACACCTGACCGCTTCTGACCTGGGTACTGTGCTC 1445
Db 1558 ccagctagacaagaacacccctagacacccctgacgcctcttacccttggtaacctgtctc 1617
QY 1446 CCTGACCGCCGAGAGATGAGTACCTCCGTGCCCCCAGACAGCATCTGGGGGTACGGCCCCA 1505
Db 1618 cctcagccccagagagctcgagctccgtgcccccaagcagcctcgtggtcaggtccccc 1677
QY 1506 GGACCTGAGACAGTGTGACCCCAAGGACAGTGGAGTCTCTATCCAGAGCGCCGCTTGC 1565
Db 1678 ggacctgagacacgtgtgacccaagcagcttgagcgtctctatcccaagggccgcttgc 1737
QY 1566 TTTTCAGAAATGAAAGGGTCCGAATCTTGTGTAAGATTCAGTCTTCTGAGTGGGGC 1625
Db 1738 ttccagaaacatgaacggtctcgaatactcgtgaaagatccagctcttccgtgtggggc 1797
QY 1626 CCCACGAGAGATTTGAAGGGGCTCAGTACAGAAATGTGAGCAATGAGCACTTGGCCAGCTT 1685
Db 1798 cccacaggaagatctgaagagcgtcagtcagaaatgtgagcatgagccttgccacgct 1857
QY 1686 CATGAAGCTGGGAGCGGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1745
Db 1858 catgaagctggagcaggtgtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1917
QY 1746 GGGACCCACCTGGAAGGCTTGAAGGCGGAGAGCGGACCGCCGCTGCGGAGTGGAT 1805
Db 1918 gggacccacatgagggcttgaagggcctgaagggcagcagcccggtgtggtgtgtgtgt 1977
QY 1806 CCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1865
Db 1978 cctagggagggagggagggagggagggagggagggagggagggagggagggagggagggag 2037
QY 1866 CAACGGCTACTGTGCTCAGACCTGACGCTGCAAGAGACCTCTCGGGAGGAGCCCTGCT 1925
Db 2038 caacggctactgtgtctcagacctgacgctgcaagagccctctcgtggagagccctgtgct 2097
QY 1926 CCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1985
Db 2098 cctagggagggagggagggagggagggagggagggagggagggagggagggagggagggag 2157
QY 1986 AGGCGCCACATCCCTTGTGCTGAGCCAGCCCTGCTGAGGATCCCGCTGGGAGAGGAG 2045
Db 2158 aggcgcacaccccttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2217
QY 2046 GCACGGGTGATCCCGTTCACCCCAAGAGAACTGCGCTAGTAAACGGAGCAATGCC 2105
Db 2218 gcacgggtgatcccggttcaccccaagagaactcgcgtcagtaaacggagacatgccc 2277
QY 2106 CCTGCAGACA 2115
Db 2278 cctgcagaca 2287
```

## RESULT 13

US-09-710-286-2845

Sequence 2845, Application US/09710286

GENERAL INFORMATION:

APPLICANT: Gearling, David P.

APPLICANT: McCarthy, Sean A.

APPLICANT: Holtzman, Douglas A.

TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES

FILE REFERENCE: 1600.2005-001

CURRENT APPLICATION NUMBER: US/09/710.286

PRIOR APPLICATION NUMBER: 60/164.255

PRIOR FILING DATE: 1999-11-09

NUMBER OF SEQ ID NOS: 4115

SOFTWARE: FastSeq for Windows Version 4.0

```
SEQ ID NO 2845
LENGTH: 2289
TYPE: DNA
ORGANISM: Homo sapiens
US-09-710-286-2845
```

Query Match 92.5% Score 1976.8; DB 28; Length 2289;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 2053; Conservative 0; Mismatches 27; Indels 30; Gaps 3;

```
QY 12 TGGCGGCGACATCCCGTGTGCTGATGAGCGGCGAGACAGACATACCGGTGAGCCACAGT 71
Db 202 tggcggcgacatcccggtgtgctgagcggtgagcgagagagagctacagcttgagccacagct 261
QY 72 GCTTCCTTCCTGGAGATACACAGACCATGAGCCCTT-CAACGGCTGACCCCTGTTGCT 130
Db 262 gcttccttcctggagatcacagacccatgagccatgagccatgagccatgagccatgagcc 321
QY 131 CCGTGGGGGAGCGGCC-----TGGCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 185
Db 322 gtcctgtggaaccccgccctcgagcctcctgtctcctgtctcctcctcagcctcagatggt 381
QY 186 GCATCCCGCGAGACCCCTGCGCTGAGAGACAGGACGAGTCTGCCCCCTGGGGGAGT 245
Db 382 gcagccctcgagacccctgctgtgagagacagagagagagagctgagccctcgagagagct 441
QY 246 CCGTGAACACCCCGCTTACCATTTTCCAGCTCTCCCTGCGCAACTCTCTGCTGCTGCTG 305
Db 442 cctgtgcaacaccccgcttaccatcttccagcctcctcctcagcctcctcctcctcctcct 501
QY 306 TGGGAGAGTCTCGGAGCTGAGCAGGAGCTGTCTCGGAGGCTGAGCTGAGCTGAGCTGAGCA 365
Db 502 tgggagagtctcgagcctgagcagagagagagagagagagagagagagagagagagagagag 561
QY 366 GAAGATGTCAAGCTCTGACAGAGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAGCC 425
Db 562 gaagatgtcaagctctgacagagagagagagagagagagagagagagagagagagagagag 621
QY 426 CCGCGAGAGCTGAGAGGCGCTCCCATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 485
Db 622 cccgagagacctgagagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 681
QY 486 CTGCGGCGCCAGAGGCTGACCCGCTTCTCTCCGCAATCAGAAAGGCAATGTGAGACT 545
Db 682 ctgcygggccccagagcctgacacgcttctctcccgatcacagagagagagagagagagag 741
QY 546 GCTCCGAGGGGGGCTCCGAGCGACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 605
Db 742 gctccgagggggtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 801
QY 606 TGTGGGGGGGCTGCTGCTGAGCGAGGCTGATGCTCGGGCTGTGGAGGCGCTGCTGAGCA 665
Db 802 tgtcggggggtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 861
QY 666 CCGTGGGGGCGCTTGTGCGCGAGTCCGCGAAGTGTCTACCCCGGCTGTGAGAGTGTG 725
Db 862 cctgctgtggcgcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 921
QY 726 CCGGAGACCCCTGAGACAGAGACAGAGAGGACAGGCGGCTGTGCAAGGCGGGGG 785
Db 922 ccggagacccctgagacagagacagagagagagagagagagagagagagagagagagag 981
QY 786 ACCCGCCAGAGGCGCCCGGTGAGATGCTGCTCCAGAGTGGAGCGCTGCGGGGGCT 845
Db 982 accccctcagggcccccggtgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1041
QY 846 GCTGCGCGTGTGAGGAGCCCATCATCCGAGCATCCCGAGGAGCATCATGTGCGCGGTG 905
Db 1042 gctgccccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1101
QY 906 GCGGCAACGCTCTCTGTGGGAGCCATCTGCGGAGGCTTGAAGGAGCATCTCTCGGCG 965
```

[illegible]

OY	2046	GCACGGGTGATCCCCCTTCCACGCCCAAGACACTCGGGCTCAGTAAGGAAGGAAATCATCC	2105
Db	2218	gcacgggtgattcccccttccaccgaagaactcgcgtcaagtaaacggaaatgccc	2277
OY	2106	CCTGCAGACA	2115
Db	2278	cctgcagaca	2287
 RESULT 14 US-09-726-171-1716 Sequence 1716, Application US/09726171 GENERAL INFORMATION: APPLICANT: Geating, David P. APPLICANT: Holtzman, Douglas A. TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES TITLE OF INVENTION: THEREFOR FILE REFERENCE: 1600.2015-001 CURRENT APPLICATION NUMBER: US/09/726, 171 CURRENT FILING DATE: 2000-11-29 PRIOR APPLICATION NUMBER: 60/167,845 PRIOR FILING DATE: 1999-11-29 NUMBER OF SEQ ID NOS: 2579 SOFTWARE: FastSeq for Windows Version 4.0 SEQ ID NO 1716 LENGTH: 2289 TYPE: DNA ORGANISM: Homo sapiens US-09-726-171-1716			
 Query Match            92.5%; Score 1976.8; DB 29; Length 2289; Best Local Similarity    97.3%; Pred. No. 0; Matches 2053; Conservative    0; Mismatches    27; Indels    30; Gaps    3;			
OY	12	TGGCGGCGACTCCCGTGCTGTGACGCGCGAGACAGACTACCGGTGAGCCACAGGT	71
Db	202	tggcggcgactcccgctgtgtgacgcgcgcgacagacctaccggtagccccaaggt	261
OY	72	GCGTCCTTCCTGGGATTTACACAGACCATTGGCTTG--CAACGGCTGACCCCTGTGGT	130
Db	262	gccttccttccttggaatatcacagaccatgagccttcgcaaggctcgacccctgttggg	321
OY	131	CCTGGGGGAGCGCCC-----TGGGAGCGCTCTGTCTCGCTTTCACGCTCGGATGGGT	185
Db	322	gtctcgttggagaccccgccctcggagagctctctgtctcttcagccctcgatagtgt	381
OY	186	GCATCCCGCGAGAGACCTGTGCTGGAGACAGAGCACGGAGTCTGCCCCCTGGGGGAGT	245
Db	382	gcagccctcggagaccctcgctggagagacagggagagagctcgccccctggagaggt	441
OY	246	CCTGCACACCCCCCATTAATTCCAGCTCTGCCCTGCGCAACTCTTGGCTTCCGCTG	305
Db	442	cttgcgccaacccaactaacatltccagccttcccctgcgaaccccttgcctccgvtg	501
OY	306	TGCGGAGGTGTCCGCTCTGAGACAGGAGACCTGTCCGCGAGGCTGGCTGTGGCTTGGACA	365
Db	502	tgcggaggtgtccgcgcctcggagcagagcagjctgcggagagctgtgctgtgccttggaca	561
OY	366	GAGAATGTCAAAGTCTTAACAGAGACGACTGGCGTGTGTGGCTTACCGGCTCTGAAGC	425
Db	562	gaagaattcaagaactctaacaagajcagctgcgtctgtgtctcaacgcgcctctgaagc	621
OY	426	CCCCGAGGACCGTGGAGCGGCTCCCAATTTGGACTCGCTGTATTTCACACCCAGATGACTT	485
Db	622	ccccgaagaccttggagcgccctcccatgtgacctctgtctatactcaaccaagatgggtc	681
OY	486	CTCGGGGGCCCAAGCGCTCACCCCTTTCTTWTCTCCGCAATCAGAAGGCCAAATGTGACCT	545
Db	682	ctcggggccccagagcgtgacocgcttctctctccgcgtcaacgaagsgcaatgtgacct	741
OY	546	GCTCCCGAAGGGGGCTCCCGAGACGACACGCTGTGTGGCTCGGGCTCTGCGCTGTGGGG	605





**THIS PAGE BLANK (USPTO)**



QY	12	TGCGCGGCACTCCGCTGTGTGGTC-GCGCGACAGAGAGATACCGGTGGACCCACGG	70
Db	44	Tggcgcggcactccgcgtctgtctgtacagcgcgcagagagatcacgcgtgacccaccgcg	103
QY	71	TGCCCTCCTCCTGGGATCTACACAGACATGGGCTTG-CACGGCTGCAGCCCTGTGG	128
Db	104	tgcctccctccctggtgatactacagacacatgacctgtgcacagcgctcgcacctgttg	165
QY	130	TCCTGTGGGACCGCC----TGGCAGCCTCTGTTCCTGCTCTTACGCTCGGATGG	184
Db	164	gtacctgtgggaccccccgcctcgcgcagcctccctgctctcttcacgctcgatgg	221
QY	185	TGCATCCCGGCAAGACCTTGCGTGGAGAGAGAGACGAGATCTGCCCCCTGGGGGAG	244
Db	224	tgcagcctctgagagacctgtgctgagagagacagcagatgctgcacccctctgacgag	283



```
Db 249 attccagcctctccctccgcaacctcttgctccctcgtgctgagggctgctgagcctg 308
QY 325 AGCAGGAGAGGTGTCCGGAGAGCTGGCTGTGGCTTGGCAGAGAAGATGTCAGCTCTCA 384
Db 309 agcacggagcgtgtccggagcgtgctgctgcttgacagagaagtgtcaagctctca 368
QY 385 ACAGAGAGAGCTGGCTGTGGCTCACCGGCTCTGAGCCCGCCGAGAGCTGAGCGCC 444
Db 369 acagagagcgtgctgctgctgctcaccggtctcagcccccagagacgtgagcgc 428
QY 445 CTCGCATTGGACC 457
Db 429 ctcccatggacc 441
```

```
RESULT 4
US-09-970-966-126
; Sequence 126, Application US/09970966
; GENERAL INFORMATION:
```

```
; APPLICANT: Stolck, John A.
; APPLICANT: Mollesh, David Alan
; APPLICANT: Eling, Steven P.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
; CURRENT FILING DATE: 2001-10-02
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 126
```

```
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-970-966-126
```

```
Query Match 10.3%; Score 219.8; DB 5; Length 396;
Best Local Similarity 80.9%; Pred. No. 3,5e-38;
Matches 279; Conservative 0; Mismatches 42; Indels 24; Gaps 1;
```

```
QY 1243 ATTGCAAGTGTGAGTGTGACGTCCCTGAGAGACCTGTAAGCCTTGTGTAAGTGCAGCAA 1302
Db 8 actcgcaagtgtgagctgacgtccctgagagacctgagccttgctgagtaacacaa 67
QY 1303 GGGCAGCAATGAGTCTCAGGCTCTCGGGGCGCCCTCCACAGGTGGCCACCTGATC 1362
Db 68 gggcagcaaatgagctc-----caggtggccacctgac 103
QY 1363 GACCGCTTTGGAAGGAGGAGGCGCAGTACAAAGACACCCCTAGACACCTGACCGCC 1422
Db 104 gacgcctctgtaaggaagggagcagctagacaagaccctagacacctgacgcgc 163
QY 1423 TTTCACCTGGGTACCTGCTCCCTCAGCCCGCAGAGAGTGAAGTCTGCTGCCCGCAG 1482
Db 164 ttctaaccttgtaacctgctcctcctcagcccgagagagctgagctcgtgcccccaagc 223
QY 1483 AGCATCTGGGCGGTGAGGCGCCAGAGACCTGAGACAGTGTACCAAGCAGCAGCTGACGTC 1542
Db 224 agcatctggcggtgtcagagcccgacagctgagacagctgagctgtaacagggcgcg 283
QY 1543 CTGTATCCCAAGGCGCCCTTGTCTTCCAGAACATGAACGGGTCC 1587
Db 284 atcccaagggctacctgctcctcagacctcagcatgcaagagcgc 328
```

```
RESULT 5
US-60-325-795-1
; Sequence 1, Application US/60325795
; GENERAL INFORMATION:
; APPLICANT: Su, Eric W
; APPLICANT: Wang, He
```

```
; APPLICANT: Zhi, Yu
; TITLE OF INVENTION: NOVEL SECRETED PROTEINS AND THEIR USES
; FILE REFERENCE: P-14990
; CURRENT APPLICATION NUMBER: US/60/325,795
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2511
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)..(2306)
; OTHER INFORMATION:
US-60-325-795-1
```

```
Query Match 2.5%; Score 52.8; DB 6; Length 2511;
Best Local Similarity 50.6%; Pred. No. 0.073;
Matches 157; Conservative 0; Mismatches 147; Indels 6; Gaps 1;
```

```
QY 504 CACCGTTTCTTCTCCGATCAGAGAGCCCAATGTGACCTGCTCCGAGGGGGCTCC 563
Db 860 caccctgacctcgcgggagcaaccagctagacaccttgccccgctgagggccgcga 919
QY 564 CGAGCAGACGCGCTGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCTGCT 623
Db 920 gctgagcgagctgagctgagctgagctgagctgagctgagctgagctgagctgagc 979
QY 624 GAGCAGAGCTGATGTGCGGCTGAGGAGGCTG-----GTTGCAAGCTGCTGAGCG 677
Db 980 actgagatgctgctgagcgagcgagctgagctgagcgagcgagcgagcgagcgagcg 1039
QY 678 CTTTGTCGCGAGTGGCGCGAGAGTGTCTACCTCCGCTGCTGAGCTGCGCGAGCCCT 737
Db 1040 cctgagggagctgagctgagctgagctgagctgagctgagctgagctgagctgagc 1099
QY 738 GAGCAGAGCAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 797
Db 1100 ggcgaggaagaaagagctggaagagcgagctgagcgagcgagcgagcgagcgagc 1159
QY 798 CCCCCCGTCG 807
Db 1160 cggccctcgcg 1169
```

```
RESULT 6
US-09-778-927A-16
; Sequence 16, Application US/09778927A
; GENERAL INFORMATION:
; APPLICANT: KHOSRAVI, Rami et al.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL
; FILE REFERENCE: 2786-0160P
; CURRENT APPLICATION NUMBER: US/09/778,927A
; CURRENT FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: ID 134453
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: IL135341
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 4282
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(4282)
; OTHER INFORMATION: n = a,c,g,t any unknown or other
US-09-778-927A-16
```



```

: Sequence 2456, Application US/09760446A
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PTZ68
: CURRENT APPLICATION NUMBER: US/09/760,446A
: PRIOR FILING DATE: 2000-01-16
: PRIOR APPLICATION NUMBER: 60/179,065
: PRIOR FILING DATE: 2000-01-31
: PRIOR APPLICATION NUMBER: 60/180,628
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: 60/214,886
: PRIOR FILING DATE: 2000-06-28
: PRIOR APPLICATION NUMBER: 60/217,487
: PRIOR FILING DATE: 2000-07-11
: PRIOR APPLICATION NUMBER: 60/225,758
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/220,963
: PRIOR FILING DATE: 2000-07-26
: PRIOR APPLICATION NUMBER: 60/217,496
: PRIOR FILING DATE: 2000-07-11
: PRIOR APPLICATION NUMBER: 60/225,447
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/218,290
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: 60/225,757
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/226,868
: PRIOR FILING DATE: 2000-08-22
: PRIOR APPLICATION NUMBER: 60/216,647
: PRIOR FILING DATE: 2000-07-07
: PRIOR APPLICATION NUMBER: 60/225,267
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/216,880
: PRIOR FILING DATE: 2000-07-07
: PRIOR APPLICATION NUMBER: 60/225,270
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/251,869
: PRIOR FILING DATE: 2000-12-08
: PRIOR APPLICATION NUMBER: 60/235,834
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: 60/234,274
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: 60/234,223
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: 60/228,924
: PRIOR FILING DATE: 2000-08-30
: PRIOR APPLICATION NUMBER: 60/224,518
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/236,369
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: 60/224,519
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/220,964
: PRIOR FILING DATE: 2000-07-26
: PRIOR APPLICATION NUMBER: 60/241,809
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/249,299
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: 60/236,327
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: 60/241,785
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/244,617
: PRIOR FILING DATE: 2000-11-01
: PRIOR APPLICATION NUMBER: 60/225,268
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/236,368
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: 60/251,856
: PRIOR FILING DATE: 2000-12-08
: PRIOR APPLICATION NUMBER: 60/251,868
: PRIOR FILING DATE: 2000-12-08
: PRIOR APPLICATION NUMBER: 60/229,344
: PRIOR FILING DATE: 2000-09-01
: PRIOR APPLICATION NUMBER: 60/234,997
: PRIOR FILING DATE: 2000-09-25
: PRIOR APPLICATION NUMBER: 60/229,343
: PRIOR FILING DATE: 2000-09-01
: PRIOR APPLICATION NUMBER: 60/229,345
: PRIOR FILING DATE: 2000-09-01
: PRIOR APPLICATION NUMBER: 60/229,287
: PRIOR FILING DATE: 2000-09-01
: PRIOR APPLICATION NUMBER: 60/229,513
: PRIOR FILING DATE: 2000-09-05
: PRIOR APPLICATION NUMBER: 60/231,413
: PRIOR FILING DATE: 2000-09-08
: PRIOR APPLICATION NUMBER: 60/229,509
: PRIOR FILING DATE: 2000-09-05
: PRIOR APPLICATION NUMBER: 60/236,367
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: 60/237,039
: PRIOR FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 60/237,038
: PRIOR FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 60/236,370
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: 60/236,802
: PRIOR FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 60/237,037
: PRIOR FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 60/237,040
: PRIOR FILING DATE: 2000-10-02
: PRIOR APPLICATION NUMBER: 60/240,960
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/239,935
: PRIOR FILING DATE: 2000-10-13
: PRIOR APPLICATION NUMBER: 60/239,937
: PRIOR FILING DATE: 2000-10-13
: PRIOR APPLICATION NUMBER: 60/241,787
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/246,474
: PRIOR FILING DATE: 2000-11-08
: PRIOR APPLICATION NUMBER: 60/246,532
: PRIOR FILING DATE: 2000-11-08
: PRIOR APPLICATION NUMBER: 60/249,216
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: 60/249,210
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: 60/226,681
: PRIOR FILING DATE: 2000-08-22
: PRIOR APPLICATION NUMBER: 60/225,759
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/225,213
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/227,182
: PRIOR FILING DATE: 2000-08-22
: PRIOR APPLICATION NUMBER: 60/225,214
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/235,836
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: 60/230,438
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/215,135
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: 60/225,266
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/249,218
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: 60/249,208
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: 60/249,213
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: 60/249,212
: PRIOR FILING DATE: 2000-11-17
: PRIOR APPLICATION NUMBER: 60/249,207

```

PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/232,400  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/231,242  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,081  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,080  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,414  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,244  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233,064  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/233,063  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,397  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,399  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,401  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/241,808  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,826  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,786  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,221  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,475  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/231,243  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233,065  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,398

Query Match 2.3%, Score 49; DB 5; Length 11720;  
Best Local Similarity 54.8%; Pred. No. 0.84;  
Matches 97; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 598 TCGTGGGCTGCGGGGCTCTCTGTCGACGAGGCTGATGTCGGGCTCTCTGGAGGCTCTG 657  
DB 9537 TGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9596  
QY 658 GCTTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717  
DB 9537 GTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 9656  
QY 718 GTGAGCTCCCGGAGCCCTGAGCCAGAGCCAGAGCCAGAGCCAGAGCCAGAGCCAGAGCC 774  
DB 9657 GGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 9713

RESULT 10

US-09-815-264-72273/c  
Sequence 72273, Application US/09815264  
GENERAL INFORMATION:  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Cao, Yongwei  
APPLICANT: Dolson, Stanton B.  
APPLICANT: Koshi, Jeffrey M.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof  
FILE REFERENCE: 38-21(51237)G  
CURRENT APPLICATION NUMBER: US/09/815,264  
CURRENT FILING DATE: 2001-03-23  
PRIOR APPLICATION NUMBER: US 09/620,392  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: US 09/702,134  
PRIOR FILING DATE: 2000-10-31  
NUMBER OF SEQ ID NOS: 109669  
SEQ ID NO 72273  
LENGTH: 17403  
TYPE: DNA  
ORGANISM: Oryza sativa  
US-09-815-264-72273

Query Match 2.3%, Score 48.6; DB 5; Length 17403;  
Best Local Similarity 45.4%; Pred. No. 1.2;  
Matches 174; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 587 CGGCTGTGCGCTGTGCGGCTGTGCGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 646  
DB 9101 CGTGTAGGGGCGAGGACACTGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 9042  
QY 647 TGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 706  
DB 9041 CGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8982  
QY 707 TACCCCGCTGTGTGAGCTGCGCGGACCCCTGACGACGACGACGACGACGACGACGACG 766  
DB 8981 TCACGCTGTGTGTGAGCTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8922  
QY 767 CGGCTGTGAGGCGCGGAGGACCCCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826  
DB 8921 AGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8862  
QY 827 TGGAGGCTGTGCGGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 886  
DB 8861 CGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8802  
QY 887 AGGCGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946  
DB 8801 GTGAGCGCGCGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8742  
QY 947 AACGAGCATCTCTCGCGCGCGG 969  
DB 8741 CCGCGAGCGAGCGAGCGCGCGG 8719

RESULT 11  
US-09-815-264-76667  
Sequence 76667, Application US/09815264  
GENERAL INFORMATION:  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Cao, Yongwei  
APPLICANT: Dolson, Stanton B.  
APPLICANT: Koshi, Jeffrey M.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: McIninch, James  
APPLICANT: Wu, Wei  
TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof

```

: FILE REFERENCE: 38-21(51237)G
: CURRENT APPLICATION NUMBER: US/09/815,264
: CURRENT FILING DATE: 2001-03-23
: PRIOR APPLICATION NUMBER: US 09/620,392
: PRIOR FILING DATE: 2000-07-19
: PRIOR APPLICATION NUMBER: US 09/702,134
: PRIOR FILING DATE: 2000-10-31
: NUMBER OF SEQ ID NOS: 103669
: SEQ ID NO 76667
: LENGTH: 50229
: TYPE: DNA
: ORGANISM: Oryza sativa
: US-09-815-264-76667

```

	Query Match	2.2%	Score 47.8:	DB 5,	length 50229;
	Best Local Similarity	50.7%;	Pred. 2.6:	Mismatches 112;	Indels 0;
	Matches 115;	Conservative 0;		Gaps 0	
QY	1756 GTGAGGGCCCTGTGAAGCGCAGAGAGCGGCACGCCCGGTGCGGAGACTGATCTTCAAGCGCAG	1815			
Db	7150 gtygtgaggaagaagaaagagagacgcgcgtcgcggcgctgcggcgcttcgaagcaag	7209			
QY	1816 CGGCGAAGAACGACTGTGCACACAGCTGGGGGCTGTGGGGCTTACAAGGGGGCATCCCCAACGGGCTAC	1875			
Db	7210 gaggcgcggggcgatcgtgagacgcgcgttgatgttacaagcaaccggaacacgcgcgcgcgcgcgc	7269			
QY	1876 CTGGCTCTACAGACTCTTCAGGCGTGCCAMAAGACCCTCTGTGGGGACGCCCTGTGCTCTTCAGGAACT	1935			
Db	7270 taattccagtaactcctcctaagggttagagcgcgagacgfyagagctgtcctctcgcgcgcgcgcgc	7329			
QY	1936 GGACTGTATTCTCACGCTCTGGCACTAGCTCTTCAAGCCTCTCCACCTGGC	1982			
Db	7330 ggagagcgtctcatcacgcgcgctctccgcgctctccgctctccccaagcgc	7376			

```

RESULT 12
US-09-815-264-76673
; Sequence 76673, Application US/09815264
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Koshi, Jeffrey M.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingtong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Mei
; TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51237)G
; CURRENT APPLICATION NUMBER: US/09/815,264
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/620,392
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/702,134
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 109669
; SEQ ID NO 76673
; LENGTH: 76304
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(76304)
; OTHER INFORMATION: unsure at all n locations
US-09-815-264-76673

```

Query Match	2.2%	Score 47.8;	DB 5;	Length 76304;
Best Local Similarity	50.7%;	Pred. No. 3;		
Matches 115;	Conservative 0;	Mismatches 112;	Indels 0;	Gaps 0;
0y	1756 GTGAGAGGCGCTGAAAGGCGAGAGAGCGGACCCCGGCTCGGAGACTGCATCTTACGGCAG	1815		

[illegible]

```

RESULT 13
US-09-815-264-80624
; Sequence 80624, Application US/09815264
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Cao, Yongwei
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Koshi, Jeffrey M.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: McIninch, James
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Genomic Plant Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51237)G
; CURRENT APPLICATION NUMBER: US/09/815,264
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/620,392
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/702,134
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 109669
; SEQ ID NO 80624
; LENGTH: 9648
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(9648)
; OTHER INFORMATION: unsure at all n locations
US-09-815-264-80624

```

[illegible]







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2001, 10:02:42 ; Search time 1369.83 Seconds  
(without alignments)  
16771.771 Million cell updates/sec

Title: US-09-215-035-1  
Perfect score: 2138  
Sequence: 1 AGGATTCGCGGCCGCC.....AAAAAAAAAAAAAAAAAAAA 2138

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estda:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hlc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hlc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vtl:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	946.4	44.3	1037	10	AL574127
C 2	903.6	42.3	1011	10	AL571138
C 3	898.6	42.0	989	10	AL578994
C 4	890.6	41.7	1051	10	AL574128
C 5	873.2	40.8	923	10	AL556628
C 6	831.8	38.9	933	10	AL578453
C 7	828.6	38.8	968	10	AL577458
C 8	821.4	38.4	953	10	AL578289
C 9	815.4	38.1	965	10	AL556862
C 10	782	36.6	912	10	AL514070
C 11	707	33.1	726	10	AL556902
C 12	677.8	31.7	817	10	AL555912

C 13	677	31.7	737	10	AI813749	AI813749	wk79a05.x
C 14	650.2	30.4	759	10	AI984029	AI984029	w52e09.x
C 15	643.8	30.1	799	10	AI744747	AI744747	tr14b11.x
C 16	641.4	30.0	825	10	AL554741	AL554741	AL554741
C 17	637	29.8	747	10	AW083412	AW083412	xc14a01.x
C 18	614.4	28.7	704	10	BE379303	BE379303	601238040
C 19	613.2	28.7	731	10	AI744989	AI744989	tr18c01.x
C 20	605.8	28.3	678	10	AI744296	AI744296	tr08h06.x
C 21	599.2	28.0	780	11	BE903130	BE903130	601673130
C 22	599.2	28.0	788	11	BC251659	BC251659	602363552
C 23	599	28.0	699	10	AL556126	AL556126	AL556126
C 24	567.2	26.5	740	10	AW105388	AW105388	xc61c03.x
C 25	566.4	26.5	992	10	AL545401	AL545401	AL545401
C 26	562.4	26.3	699	10	AI587660	AI587660	tr80d05.x
C 27	541.2	25.3	610	11	BI333483	BI333483	602996934
C 28	537.6	25.1	690	10	AI680542	AI680542	tr82c10.x
C 29	537.4	25.1	724	11	BE900825	BE900825	601674556
C 30	531.2	24.8	537	10	AA488406	AA488406	ad37g03.s
C 31	528.8	24.7	586	10	AW272314	AW272314	xu17f12.x
C 32	519.2	24.3	572	10	AI580209	AI580209	LI93c06.x
C 33	518.4	24.2	557	10	AW518996	AW518996	ha47h03.x
C 34	499.4	23.4	509	10	AI859791	AI859791	wm20e02.x
C 35	498.2	23.3	503	10	AW474904	AW474904	xy21d10.x
C 36	492.8	23.0	568	10	AW102974	AW102974	xd41c09.x
C 37	489.8	22.9	610	10	AW339085	AW339085	ha71a09.x
C 38	482.2	22.6	487	10	BE271658	BE271658	601141233
C 39	482.2	22.6	487	10	AI567987	AI567987	tr86g11.x
C 40	479.6	22.4	513	10	AI095455	AI095455	cb19g04.x
C 41	478	22.4	486	10	AI342440	AI342440	q139c10.x
C 42	476.8	22.3	503	10	AI612761	AI612761	tp4a811.x
C 43	462.8	21.6	542	10	AL556629	AL556629	AL556629
C 44	461.8	21.6	916	11	BG335095	BG335095	602403573
C 45	441.2	20.6	446	10	BE713092	BE713092	IL5-HT070

## ALIGNMENTS

RESULT 1  
LOCUS AL574127/c  
DEFINITION AL574127 LTI\_NFL006.PL2 Homo sapiens cDNA clone CSOD1040YH04 3  
ACCESSION AL574127  
VERSION AL574127.1 GI:12934032  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1037)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.

## FEATURES

source

1. 1037  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSOD1040YH04"  
/clone\_id="LTI\_NFL006.PL2"  
/issue="Vector: PCWSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com"  
 BASE COUNT 182 a 313 c 332 g 208 t 2 others  
 ORIGIN

Query Match 44.3%; Score 946.4; DB 10; Length 1037;  
 Best Local Similarity 94.5%; Pred. No. 1.4e-135;  
 Matches 1032; Conservative 2; Mismatches 3; Indels 55; Gaps 3;

QY 953 CCATCTCCGCGCGGTTCCGCGGGAAGTGAGAGACACCTGCTCTTACGACAGA 1012  
 Db 1037 CCMTCCTCCGGCGCGGTTCCGCGGGAAGTGAGAGACACCTGCTCTTACGACAGA 978  
 QY 1013 AGGCGCGGAGATGAGAGAGAGCCCTCATCTCTTACAGAGGGAGGTGAGAGCTGCG 1072  
 Db 977 AGGCGCGGAGATGAGAGAGAGCCCTCATCTCTTACAGAGGGAGGTGAGAGCTGCG 918  
 QY 1073 TGGATGCGGCGCTGCTGCGCAACAGATGAGACCGCGTGAAGCCATCCCTTACACTAG 1132  
 Db 917 TGGATGCGGCGCTGCTGCGCAACAGATGAGACCGCGTGAAGCCATCCCTTACACTAG 858  
 QY 1133 AGCAGCTGGAGCTCTTAAGATTAACCTGATGAGCTTACCCACAGATTACCCGAGT 1192  
 Db 857 AGCAGCTGGAGCTCTTAAGATTAACCTGATGAGCTTACCCACAGATTACCCGAGT 798  
 QY 1193 CTGTGATCAGACACCTGCGGCTACCTCTCTCAAGATGAGCCCTGAGGACATTCGCAAGT 1252  
 Db 797 CTGTGATCAGACACCTGCGGCTACCTCTCTCAAGATGAGCCCTGAGGACATTCGCAAGT 738  
 QY 1253 GGAATGTACGTCCTCTGAGACACCTGGAAGGCTTGTGAAGTCAACAAAGGCGACGAAA 1312  
 Db 737 GGAATGTACGTCCTCTGAGACACCTGGAAGGCTTGTGAAGTCAACAAAGGCGACGAAA 678  
 QY 1313 TGAGTCTCAGAGCTCTCTGCGGCGCCCTCCACAGTGGGCCACCTGATTCGACCGCTTGT 1372  
 Db 677 TGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 642  
 QY 1373 TGAAGGAGAGGCGGAGCTAGACAAAGACACCTTACACCTGACCTCTTCTACCTG 1432  
 Db 641 TGAAGGAGAGGCGGAGCTAGACAAAGACACCTTACACCTGACCTCTTCTACCTG 582  
 QY 1433 GGTACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1492  
 Db 581 GGTACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 522  
 QY 1493 CGGTCAGGCGGCGGAGCTTGAACAAGTGAACCGGAGCTGAGAGCTGCTCTATCCCA 1552  
 Db 521 CGGTCAGGCGGCGGAGCTTGAACAAGTGAACCGGAGCTGAGAGCTGCTCTATCCCA 462  
 QY 1553 AGGCGCGGCTTGTCTTCCAGAACATGAACGGGTCGGAATCTTCTGTAAGATCCAGTCT 1612  
 Db 461 AGGCGCGGCTTGTCTTCCAGAACATGAACGGGTCGGAATCTTCTGTAAGATCCAGTCT 402  
 QY 1613 TCCGTGGTGGGCGGCGGAGGATTTGAAGGCGCTAGTCAGACAGATGTAGCATGG 1672  
 Db 401 TCCGTGGTGGGCGGCGGAGGATTTGAAGGCGCTAGTCAGACAGATGTAGCATGG 342  
 QY 1673 ACTTGGGCGGCTTGTCTTCCAGAACATGAACGGGTCGGAATCTTCTGTAAGATCCAGTCT 1732  
 Db 341 ACTTGGGCGGCTTGTCTTCCAGAACATGAACGGGTCGGAATCTTCTGTAAGATCCAGTCT 282  
 QY 1733 TGCAGAAACTTCTGAGAGCCCAAGTGAAGGCTGAAGGCGGAGAGGCGGACCGCGCGG 1792  
 Db 281 TGCAGAAACTTCTGAGAGCCCAAGTGAAGGCTGAAGGCGGAGAGGCGGACCGCGCGG 222  
 QY 1793 TGCAGAAACTTCTGAGAGCCCAAGTGAAGGCTGAAGGCGGAGAGGCGGACCGCGCGG 1852  
 Db 221 TGCAGAAACTTCTGAGAGCCCAAGTGAAGGCTGAAGGCGGAGAGGCGGACCGCGCGG 163  
 QY 1853 AGGCGGCGGATCCCAAGGCTTACGAGGCTGAGAGCTTACGAGGCTGAGAGGCTTACGAGG 1912  
 Db 162 AGGCGGCGGATCCCAAGGCTTACGAGGCTGAGAGCTTACGAGGCTGAGAGGCTTACGAGG 115

QY 1913 GGAGCGGCTGCTCTTACGAGCTGAGACCTGTTCTTACCGCTGCGACTGCTTACGCT 1972  
 Db 116 -----AGGACTGAGACCTGTTCTTACCGCTGCGACTGCTTACGCT 73  
 QY 1973 CCACCTGGCTGAGGCGGCGGAGCTGCTTGTGCGGCGGAGCTGCTGAGGAGTCCCGCC 2032  
 Db 72 CCACCTGGCTGAGGCGGCGGAGCTGCTTGTGCGGCGGAGCTGCTGAGGAGTCCCGCC 13  
 QY 2033 TGGCCAGGAGCA 2044  
 Db 12 TGGCCAGGAGTCA 1

RESULT 2  
 AL571138/c  
 LOCUS AL571138 1011 bp mRNA EST 16-FEB-2001  
 DEFINITION AL571138 LTI\_NFL006.PL2 Homo sapiens cDNA clone CS0D1027YL14 3  
 ACCESSION AL571138  
 VERSION AL571138.1 GI:12928134  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1011)  
 Li, W. B., Gruber, C., Jesssee, J., and Polayes, D.  
 Full-length cDNA libraries and normalization.  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 source  
 1..1011  
 Location/Qualifiers

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1027YL14"  
 /clone.lib="LTI\_NFL006.PL2"  
 /tissue="type="placenta"  
 /note="Vector: pCMVSPORT 6. Site 1: NotI; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed by  
 Life Technologies. Contact : Feng Liang Life Technologies,  
 a division of Invitrogen 9600 Medical Center Drive  
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
 Email : filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 177 a 300 c 331 g 199 t 4 others  
 ORIGIN

Query Match 42.3%; Score 903.6; DB 10; Length 1011;  
 Best Local Similarity 96.7%; Pred. No. 4.7e-129;  
 Matches 940; Conservative 1; Mismatches 7; Indels 24; Gaps 1;

QY 1112 AGGCGATCCCTTACCTACGAGGAGCTGAGCGTCTTAACGATTAAGTATGAGCTCT 1171  
 Db 948 AGGCGATCCCTTACCTACGAGGAGCTGAGCGTCTTAACGATTAAGTATGAGCTCT 889  
 QY 1172 ACCGACAGGTTACCCGAGCTGATCGACAGCTGGCTTACCTCTTCAAGATGA 1231  
 Db 888 ACCGACAGGTTACCCGAGCTGATCGACAGCTGGCTTACCTCTTCAAGATGA 829  
 QY 1232 GCCCTGAGACATTCGCAAGTGAATGAGCTCCCTGAGACACCTGAAGGCTTGTGCTG 1291  
 Db 828 GCCCTGAGACATTCGCAAGTGAATGAGCTCCCTGAGACACCTGAAGGCTTGTGCTG 769  
 QY 1292 AAGTGCAGAAAGGCGAGAAATGAGTCTCTAGGCTCTCTGCGGCGGCTCCGACAGGCTG 1351  
 Db 1292 AAGTGCAGAAAGGCGAGAAATGAGTCTCTAGGCTCTCTGCGGCGGCTCCGACAGGCTG 115

```

Db 768 AAGTCACAAAGGCGACGAATGAGTCTT-----CAGGTGG 733
QY 1352 CCACCTGATGACCGCTTTGTGAAGGAAGGGCCAGCTAGACAAGACCCCTAGACA 1411
Db 732 CCACCTGATGACCGCTTTGTGAAGGAAGGGCCAGCTAGACAAGACCCCTAGACA 673
QY 1412 CCCGACCGCTTTGTACCCCTGGGTAGCTGTGCTCCCTCAGCCCGGAGAGCTGAGCTCG 1471
Db 672 CCGTACCGCTTTGTACCCCTGGGTAGCTGTGCTCCCTCAGCCCGGAGAGCTGAGCTCG 613
QY 1472 TGGCCCCCAGAGATCTGGGGCGGTCAAGGCCAGAGACCTGAGACAGTGTACCCCAAGG 1531
Db 612 TGGCCCCCAGAGATCTGGGGCGGTCAAGGCCAGAGACCTGAGACAGTGTACCCCAAGG 553
QY 1532 AGCTGAGAGCTCTTATCTCCCAAGGCCCGCTTGTCTTCCAGACATGAGACGGTCCGAT 1591
Db 552 AGCTGAGAGCTCTTATCTCCCAAGGCCCGCTTGTCTTCCAGACATGAGACGGTCCGAT 493
QY 1592 ACTTGTGAGAGATCCAGTCTTCTGTGGGTGGGGCCCCCAGGAGGATTTGAAGCGCTCA 1651
Db 492 ACTTGTGAGAGATCCAGTCTTCTGTGGGTGGGGCCCCCAGGAGGATTTGAAGCGCTCA 433
QY 1652 GTGAGCAAGATGTAGATGAGACTTGGCCAGACTGATGAGCTGGAGCGATCGGCTGC 1711
Db 432 GTGAGCAAGATGTAGATGAGACTTGGCCAGACTGATGAGCTGGAGCGATCGGCTGC 373
QY 1712 TGCCGTTGACTGTGTGAGGTGAGAAACTTGTGGGACCCAGCTGGAGGGGCTGAAAG 1771
Db 372 TGCCGTTGACTGTGTGAGGTGAGAAACTTGTGGGACCCAGCTGGAGGGGCTGAAAG 313
QY 1772 CGGAGAGCGGCGACCGCCGGGTGCGGGAAGTGTGATCTAAGGACGCGAGAGACGACTGG 1831
Db 312 CGGAGAGCGGCGACCGCCGGGTGCGGGAAGTGTGATCTAAGGACGCGAGAGACGACTGG 253
QY 1832 ACAGCGTGGGCGTGGGCTACAGGGCGGCGATCCCAAGCGGTACTGCTTAAGCTCA 1891
Db 252 ACAGCGTGGGCGTGGGCTACAGGGCGGCGATCCCAAGCGGTACTGCTTAAGCTCA 193
QY 1892 GCGTGAAGAGACCTCTCGGGGAGAGCCCTGCTCTAGAGACTGAGACTGTTCTCAACG 1951
Db 192 GCGTGAAGAGACCTCTCGGGGAGAGCCCTGCTCTAGAGACTGAGACTGTTCTCAACG 133
QY 1952 TCTGGAAGAGCTCTCTAGCTCCACCCCTGGGCTGAGGGGCCACTCTGCTGAGCCCA 2011
Db 132 TCTGGAAGAGCTCTCTAGCTCCACCCCTGGGCTGAGGGGCCACTCTGCTGAGCCCA 73
QY 2012 GCCCTGTGGGGATCCCGGCTGGCCAGAGACGAGACGAGTGTATCCCGCTTCAACCCA 2071
Db 72 GCCCTGTGGGGATCCCGGCTGGCCAGAGACGAGACGAGTGTATCCCGCTTCAACCCA 13
QY 2072 AGAGACTCGCG 2083
Db 12 AGAGACTCGCG 1

```

RESULT 3  
 LOCUS AL578994 989 bp mRNA EST 16-FEB-2001  
 DEFINITION AL578994 ltr1.NFL006.PL2 Homo sapiens cDNA clone CS00K012YE02 3  
 ACCESSION AL578994  
 VERSION AL578994.1 GI:12943602  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 989)  
 AUTHORS Li, W. B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage

```

FEATURES
    source
        BP 191 91006 EVRY cedex - France
        Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
        Location/Qualifiers
            1..989
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CS00K012YE02"
                /clone_1db="LTI_NF1006_PL2"
                /tissue_type="placenta"
                /note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA
                was primed with a NotI-oligo(dT) primer. Five prime end
                enriched, double-stranded cDNA was digested with Not I and
                cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                vector. Library was normalized. Library was constructed by
                life technologies. Contact : Feng Liang life technologies,
                a division of invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA fax : (1) 301 610 8371
                Email : fliang@lifetechn.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT 174 a 296 c 325 g 190 t 4 others
ORIGIN
Query Match 42.0% Score 898.6; DB 10; Length 989;
Best Local Similarity 96.1% Pred. No. 2.8e-128;
Matches 968; Conservative 4; Mismatches 8; Indels 27; Gaps 4;
QY 1063 GAACCTGCGGTGATGCGGCGCTGCTGGCCACCCAGATGAGACCGCGTAAGCCATCCCC 1122
Db 980 GAACCTGCGGTGATGCGGCGCTGCTGGCCACCCAGATGAGACCGCGTAAGCCATCCCC 922
QY 1123 TTCACTACAGAGAGCTGAGAGCTCTTAAGCATTAAGTGAAGCTTACCCAAAGT 1182
Db 921 TTCACTACAGAGAGCTGAGAGCTCTTAAGCATTAAGTGAAGCTTACCCAAAGT 862
QY 1183 TACCCGAGTGTGATCCAGACACTGGGCTAAGCTTCTCAAGATGAGCCCTGAGAG 1242
Db 861 TRCCSCGAGTGTGATCCAGACACTGGGCTAAGCTTCTCAAGATGAGCCCTGAGAG 802
QY 1243 ATTGCAAGTGAATGAGAGCTCCCTGGAGACCCCTGAAGGCTTGTGAAGTGAGACAA 1302
Db 801 ATTGCAAGTGAATGAGAGCTCCCTGGAGACCCCTGAAGGCTTGTGAAGTGAGACAA 742
QY 1303 GGGCAGCAATGATGCTCAGAGCTCTCGCGGGCCCTCCACAGAGTGGCCACCTGATC 1362
Db 741 GGGCAGCAATGATGCTC-----CAGTGGCCACCTGATC 706
QY 1363 GACCGCTTTGTGAAGGAAGGGGCCAGCTAGACAAGACACCTAGACACCTGACGCC 1422
Db 705 GACCGCTTTGTGAAGGAAGGGGCCAGCTAGACAAGACACACCTGACGCC 646
QY 1423 TTCTACCTGGGTACCTGTGCTCCCTCAGCCCGGAGAGAGTACCTCCGTCCCCCAGC 1482
Db 645 TTCTACCTGGGTACCTGTGCTCCCTCAGCCCGGAGAGAGTACCTCCGTCCCCCAGC 586
QY 1483 AGCATCTGGCGGTCAAGGCCAGAGCTGAGACAGTGTGACCCCAAGGACACTGGAGTC 1542
Db 585 AGCATCTGGCGGTCAAGGCCAGAGCTGAGACAGTGTGACCCCAAGGACACTGGAGTC 526
QY 1543 CTCTATCCCAAGGCCCGCTTGTCCAGAACATGAAGGGGTCCGATCTGTGTGAG 1602
Db 525 CTCTATCCCAAGGCCCGCTTGTCCAGAACATGAAGGGGTCCGATCTGTGTGAG 466
QY 1603 ATCCAGTCTTCTGTGGGTGGGGGCCCCAGAGAGATTTGAAGGGGCTCAGACAGAT 1662
Db 465 ATCCAGTCTTCTGTGGGTGGGGGCCCCAGAGAGATTTGAAGGGGCTCAGACAGAT 406
QY 1663 GTGAGCATGTGAGCTTGCCAGCTTATGAACTGCGAGAGATGCGGTGCTGCGCTGACT 1722
Db 405 GTGAGCATGTGAGCTTGCCAGCTTATGAACTGCGAGAGATGCGGTGCTGCGCTGACT 346
QY 1723 GTGCTGAGGTGCAAGAACTTGTGGAGCCCTGACAGGCGCTGAAGGGCGAGAGCGG 1782

```

Db 345 GTGGCTGAGTGACAGAACTTCTGGGACCCACGTGAGAGGCGCTGAAGGCGGAGGAGCGG 286  
Oy 1783 CACGCCCGGCTGGGAGTGTGATCTTACGGCAGCGGAGGAGGAGTGTGAGGAGTGTGAGG 1842  
Db 285 CACGCCCGGCTGGGAGTGTGATCTTACGGCAGCGGAGGAGGAGTGTGAGGAGTGTGAGG 227  
Oy 1843 CTGGGGCTACAGGGGCGGAGTGTGATCTTACGGCAGCGGAGGAGTGTGAGGAGTGTGAGG 1902  
Db 226 CTGGGGCTACAGGGGCGGAGTGTGATCTTACGGCAGCGGAGGAGTGTGAGGAGTGTGAGG 167  
Oy 1903 ACCCTCTGGGAGGAGGAGTGTGATCTTACGGCAGCGGAGGAGTGTGAGGAGTGTGAGG 1962  
Db 166 -GCTCTGGGAGGAGGAGTGTGATCTTACGGCAGCGGAGGAGTGTGAGGAGTGTGAGG 108  
Oy 1963 CTCTAGGCTCTACCGCTGTGAGGAGGAGTGTGATCTTACGGCAGCGGAGGAGTGTGAGG 2022  
Db 107 CTCTAGGCTCTACCGCTGTGAGGAGGAGTGTGATCTTACGGCAGCGGAGGAGTGTGAGG 48  
Oy 2023 GATCCCGGCTGTGGGAGGAGGAGTGTGATCTTACGGCAGCGGAGGAGTGTGAGG 2069  
Db 47 GATCCCGGCTGTGGGAGGAGGAGTGTGATCTTACGGCAGCGGAGGAGTGTGAGG 1

RESULT 4  
AL574128 1051 bp mRNA EST 16-FEB-2001  
LOCUS AL574128 LTI\_NF1006.PL2 Homo sapiens cDNA clone CS0D1040YH04 5  
DEFINITION prime, mRNA sequence.  
ACCESSION AL574128  
VERSION AL574128.1 GI:12934034  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1051)  
AUTHORS Li, M.B., Gruber, C., Jesse, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
Source Location/Qualifiers

1..1051  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0D1040YH04"  
/clone\_lib="LTI\_NF1006.PL2"  
/tissue\_type="placenta"  
/note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life technologies, a division of invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 153 a 364 c 341 g 187 t 6 others

Query Match 41.7% Score 890.6; DB 10; Length 1051;  
Best Local Similarity 96.1% Pred. No. 4.5e-127;  
Matches 997; Conservative 2; Mismatches 27; Indels 12; Gaps 8;

Oy 54 TACGGGTGACCCAGGCTGCTCCCTCCGATCTACAGACGACGCTTG-CAAC 112  
Db 18 TACGGGTGACCCAGGCTGCTCCCTCCGATCTACAGACGACGCTTG-CAAC 77  
Oy 113 GGCTTGACCCCTGTTGCTCTGTGGGAGCGGCC-----TGGACGCTCTCTTCTCTCT 167

Db 78 GGCTTGACCCCTGTTGGGCTCTGTGGGAGCCCGCCCTCGGACGCTCTCTCTCTCT 137  
Oy 168 CTTCAGCTCTGGATGTGATCTCCCGGAGACCTGCTGGAGAGACAGGAGGAGT 227  
Db 138 CTTCAGCTCTGGATGTGATCTCCCGGAGACCTGCTGGAGAGACAGGAGGAGT 197  
Oy 228 TGCCCGCTGGGAGGAGTCTGACAAACCCCATTAATTTTCAGACCTCTCCCTCGCA 287  
Db 198 TGCCCGCTGGGAGGAGTCTGACAAACCCCATTAATTTTCAGACCTCTCCCTCGCA 256  
Oy 288 ACTCTTGAGCTTCCGCTGTGGGAGGAGTCTGCGGCTGTGAGCAAGGAGCTGTCCGGA 347  
Db 257 ACTCTTGAGCTTCCGCTGTGGGAGGAGTCTGCGGCTGTGAGCAAGGAGCTGTCCGGA 316  
Oy 348 GGCTTGAGCTTCCGCTGTGGGAGGAGTCTGCGGCTGTGAGCAAGGAGCTGTCCGGA 407  
Db 317 GGCTTGAGCTTCCGCTGTGGGAGGAGTCTGCGGCTGTGAGCAAGGAGCTGTCCGGA 376  
Oy 408 TCACCGCTCTGTGAGCCCGGAGGAGCTGTGAGGAGCTGTCCCATTTGACCTGTCTAT 467  
Db 377 TCACCGCTCTGTGAGCCCGGAGGAGCTGTGAGGAGCTGTCCCATTTGACCTGTCTAT 436  
Oy 468 CCTACCCAGATGCTGTCTGTGGGAGGAGCTGTGAGGAGCTGTCCCATTTGACCTGTCTAT 527  
Db 437 CCTACCCAGATGCTGTCTGTGGGAGGAGCTGTGAGGAGCTGTCCCATTTGACCTGTCTAT 496  
Oy 528 GAAGGCAATGTGAGCTGTCTGTGGGAGGAGCTGTGAGGAGCTGTCCCATTTGACCTGTCTAT 586  
Db 497 GAAGGCAATGTGAGCTGTCTGTGGGAGGAGCTGTGAGGAGCTGTCCCATTTGACCTGTCTAT 556  
Oy 587 CGGCTGTGAGCTTCCGCTGTGGGAGGAGTCTGTG-CTGAGCGAGCTGTGTGAGGAGT 645  
Db 557 CGGCTGTGAGCTTCCGCTGTGGGAGGAGTCTGTG-CTGAGCGAGCTGTGTGAGGAGT 616  
Oy 646 CTGGAGGAGCTGTGGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 705  
Db 617 CTGGAGGAGCTGTGGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 676  
Oy 706 CTACCCGCTGTGAGCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 765  
Db 677 CTACCCGCTGTGAGCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 736  
Oy 766 GCGGCTGTGAGGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 825  
Db 737 GCGGCTGTGAGGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 795  
Oy 826 ATGAGGCTGTGAGGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 885  
Db 796 ATGAGGCTGTGAGGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 855  
Oy 886 CAGGAGCTGTGAGGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 945  
Db 856 CA-GMATGCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 913  
Oy 946 GAAGGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 1005  
Db 914 WAAGGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 973  
Oy 1006 GGCAAGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 1065  
Db 974 GGCAAGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 1033  
Oy 1066 GCGTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 1083  
Db 1034 GCGTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGGAGTCTGTGAGGAGT 1051

RESULT 5  
AL556628/c 923 bp mRNA EST 16-FEB-2001  
LOCUS AL556628 LTI\_NF1006.PL2 Homo sapiens cDNA clone CS0K006YH18 3  
DEFINITION prime, mRNA sequence.

```

ACCESSION AL556628
VERSION AL556628.1 GI:12899477
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 923)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
FEATURES
    source
        1..923
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_1lb="LTI_NFL006_PL2"
            /tissue_type="placenta"
            /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed by
            Lite Technologies. Contact : Feng Liang Lite Technologies,
            a division of Invitrogen 9800 Medical Center Drive
            Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
            Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"
BASE COUNT 167 a 276 c 301 g 176 t 3 others
ORIGIN
Query Match 40.8%; Score 873.2; DB 10; Length 923;
Best Local Similarity 96.5%; Pred. No. 2.1e-124;
Matches 909; Conservative 3; Mismatches 6; Indels 24; Gaps 1;
QY 1118 TCCCTTACCTGAGCAGCAGCTGAGCTCTTAAAGCATTAAGTGATGAGCTTACCCAC 1177
DB 922 TCCCTTACCTGAGCAGCAGCTGAGCTCTTAAAGCATTAAGTGATGAGCTTACCCAC 863
QY 1178 AAGGTTACCCGAGTGTGATCCAGCAGCAGCTGAGCTCTTCTCAAGTGAAGCCCTG 1237
DB 862 AAGGTTACCCGAGTGTGATCCAGCAGCAGCTGAGCTCTTCTCAAGTGAAGCCCTG 803
QY 1238 AGGACATTCGCAAGTGAATGTGACCTCTGAGAGACCTGGAAGCTTGTGCTGAAGTCG 1297
DB 802 AGGACATTCGCAAGTGAATGTGACCTCTGAGAGACCTGGAAGCTTGTGCTGAAGTCG 743
QY 1298 ACAAGGGGCAAGAAATGATGCTCAGGCTCTGCGGCGCCCTCCACAGGTGGCCACC 1357
DB 742 ACAAGGGGCAAGAAATGATGCTCAGGCTCTGCGGCGCCCTCCACAGGTGGCCACC 707
QY 1358 TGAATGACCGCTTGTGAAGGAAGGGGCAAGTGAACAAAGACACCTGAGACCCCTGA 1417
DB 706 TGAATGACCGCTTGTGAAGGAAGGGGCAAGTGAACAAAGACACCTGAGACCCCTGA 647
QY 1418 CCGGCTTCTACCTGAGTGTGCTCTCAAGCCCGGAGAGAGTGAAGTCTGCGCC 1477
DB 646 CCGGCTTCTACCTGAGTGTGCTCTCAAGCCCGGAGAGAGTGAAGTCTGCGCC 587
QY 1478 CAGACAGATCTGGGGCGTCAAGGCCCAAGACACTGAGACAGTGTGACCAAGCAGCTGG 1537
DB 586 CAGACAGATCTGGGGCGTCAAGGCCCAAGACACTGAGACAGTGTGACCAAGCAGCTGG 527
QY 1538 ACGTCTCTTACCAAGGCGCGCTTGTCTTCCAGAACATGAAGGGTCCGAATACTCG 1597
DB 526 ACGTCTCTTACCAAGGCGCGCTTGTCTTCCAGAACATGAAGGGTCCGAATACTCG 467
QY 1598 TGAAGATCAAGTCTTCTGAGTGGGGCCCAAGGAGGATTTGAAGGCGCTCAGTACG 1657

```

```

DB 466 TGAATGACAGTCTTCTGAGTGGGGCCCAAGGAGATTTGAAGCGCTCAGTACG 407
QY 1658 AGAATGAGCATGAGATTTGGCCACCTTTCATGAAGCTGCGGAGGATGCGCTG 1717
DB 406 AGAATGAGCATGAGATTTGGCCACCTTTCATGAAGCTGCGGAGGATGCGCTG 347
QY 1718 TGACTGAGGCTGAGTGCAGAACTTCTGGAGCCCAAGTGAAGGGCTGAAGCGGAG 1777
DB 346 TGACTGAGGCTGAGTGCAGAACTTCTGGAGCCCAAGTGAAGGGCTGAAGCGGAG 287
QY 1778 AGCGGACCGCCGAGTGGGAGAGCTGATCTTACGAGCGGAGAGAGTGAAGCAGC 1837
DB 286 AGCGGACCGCCGAGTGGGAGAGCTGATCTTACGAGCGGAGAGAGTGAAGCAGC 227
QY 1838 TGGGGCTGGGGCTACAGGGGGCATCCCAAGCGCTACGTGCTTACAGCTCAGCTG 1897
DB 226 TGGGGCTGGGGCTACAGGGGGCATCCCAAGCGCTACGTGCTTACAGCTCAGCTG 167
QY 1898 AAGGACCGCTTGGGGAGAGCGCCCTGCTAGAGCGGAGCTTCTACCGCTCTG 1957
DB 166 AAGGACCGCTTGGGGAGAGCGCCCTGCTAGAGCGGAGCTTCTACCGCTCTG 107
QY 1958 CACTGCTCTAGCTTCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2017
DB 106 CACTGCTCTAGCTTCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 47
QY 2018 CTGGGATTCCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2059
DB 46 CTGGGATTCCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5

```

```

RESULT 6
AL578453/c 933 bp mRNA EST 15-FEB-2001
LOCUS AL578453 LTI_NFL006_PL2 Homo sapiens cDNA clone CSDBK011YD21 3
DEFINITION AL578453
ACCESSION AL578453
VERSION AL578453.1 GI:12942540
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 933)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
FEATURES
    source
        1..933
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_1lb="LTI_NFL006_PL2"
            /tissue_type="placenta"
            /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed by
            Lite Technologies. Contact : Feng Liang Lite Technologies,
            a division of Invitrogen 9800 Medical Center Drive
            Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
            Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"
BASE COUNT 164 a 274 c 298 g 192 t 5 others
ORIGIN

```

Query Match 38.8%; Score 831.8; DB 10; Length 933;  
 Best Local Similarity 95.3%; Pred. No. 4,5e-118;  
 Matches 894; Conservative 3; Mismatches 16; Indels 25; Gaps 3;

```

QY 1123 TTCACCTACGAGCAGCTGGAGCTGCTTAAGCATTAACCTGATGAGCTTACCCACAAGT 1182
D 1123 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
D 933 TTKACCTACGAGCAGCTGGAGCTGCTTAAGCATTAACCTGATGAGCTTACCCACAAGT 874
QY 1183 TACCCGAGCTGTGATCAGCAGCTGAGCTTACCTTCTCAAGATGAGCCCTGAGAGC 1242
D 873 TACCCGAGCTGTGATCAGCAGCTGAGCTTACCTTCTCAAGATGAGCCCTGAGAGC 814
QY 1243 ATTCGCAAGTGAATGTGACGTCTCCCTGAGACCCCTGAAGGCTTTGTAAGTGCACAA 1302
D 813 ATTCGCAAGTGAATGTGACGTCTCCCTGAGACCCCTGAAGGCTTTGTAAGTGCACAA 754
QY 1303 GGGACGAAATGATGCTCCACAGCTCTCTGGGGCCCTCCACAGTGGCCACCTGATC 1362
D 753 GGGACGAAATGATGCTCCACAG-----GTGGCCACCTGATC 717
QY 1363 GACCGCTTTGTGAAGGAGGAGGCGCAGCTAGACAAAGACAC-CCTAGACACCTGACCGC 1421
D 716 GACCGCTTTGTGAAGGAGGAGGCGCAGATAGAAAAGMACACCTAGACACCTGACCGC 657
QY 1422 CTTTACCTGGGTACTGTCTGCTCCCTGAGCCCGAGAGAGCTGAGCTCCGTCCGCCAG 1481
D 656 CTTTACCTGGGTACTGTCTGCTCCCTGAGCCCGAGAGAGCTGAGCTCCGTCCGCCAG 597
QY 1482 CAGCATGCGGCGGTGACGAGGCGCCAGACCTGACACGCTGACCCCAAGGAGCTGAGAGT 1541
D 596 CAGCATGCGGCGGTGACGAGGCGCCAGACCTGACACGCTGACCCCAAGGAGCTGAGAGT 537
QY 1542 CCTATATCCCAAGGAGGCGCCTTGTCTTCCAGACATGACGAGGCTCCGAAATCTTGTGAA 1601
D 536 CCTATATCCCAAGGAGGCGCCTTGTCTTCCAGACATGACGAGGCTCCGAAATCTTGTGAA 477
QY 1602 GATCAGCTCTCTCTGGGTGGGCGCCCGCAGGAGAGATTGAAGCGCTCACTACAGACAA 1661
D 476 GATCAGCTCTCTCTGGGTGGGCGCCCGCAGGAGATTGAAGCGCTCACTACAGACAA 417
QY 1662 TGTGAGCATGAGCTTGGGCGACGTTGATGAGCTGGGAGGATGGGTGGCGCGCTGAC 1721
D 416 TGTGAGCATGAGCTTGGGCGACGTTGATGAGCTGGGAGGATGGGTGGCGCGCTGAC 357
QY 1722 TGTGAGCTGAGTGCAGAACTTCTGAGACCCACAGCTGAGGAGGCTGGAAGGCGAGAGC 1781
D 356 TGTGAGCTGAGTGCAGAACTTCTGAGACCCACAGCTGAGGAGGCTGGAAGGCGAGAGC 297
QY 1782 GCACCGCGCGGTGGGAGCTGATCTACGAGGAGGAGAGACGACGACGCTGGG 1841
D 296 GCACCGCGCGGTGGGAGCTGATCTACGAGGAGGAGAGACGACGACGCTGGG 238
QY 1842 GCTGGGGGTACAGGGGCGGATCCCAAGGAGCTTACCTGCTTACGCTCAGAGTGCAGAA 1901
D 237 GCTGGGGGTACAGGGGCGGATCCCAAGGAGCTTACCTGCTTACGCTCAGAGTGCAGAA 178
QY 1902 GACCTCTCGGGGAGGCGCTGCTCTAGAGACTGAGACTGTTCTTCAACCGCTGCGCACT 1961
D 177 GACCTCTCGGGGAGGCGCTGCTCTAGAGACTGAGACTGTTCTTCAACCGCTGCGCACT 118
QY 1962 GCTCTAGAGCTCAGCCTGAGCTGAGGAGGCGCCACTCCCTTGTGAGGCGCCAGCCTGCTGG 2021
D 117 GCTCTAGAGCTCAGCCTGAGCTGAGGAGGCGCCACTCCCTTGTGAGGCGCCAGCCTGCTGG 58
QY 2022 GGATCCCGGCTGGCGCAGAGCAGGAGGCTGATCC 2059
D 57 GGATCCCGGCTGGCGCAGAGCAGGAGGCTGATCC 20

```

RESULT 7  
 AL577458/c 968 bp mRNA EST 16-FEB-2001  
 LOCUS AL577458 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0DI086YI01 3  
 DEFINITION

prime, mRNA sequence.  
 AL577458  
 AL577458.1 GI:12940607  
 EST.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Homo sapiens  
 human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 968)  
 Li, M. B., Gruber, C., Jesse, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
 Location/Qualifiers  
 1..968  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DI086YI01"  
 /clone\_lib="LTI\_NFL006\_PL2"  
 /tissue\_type="placenta"  
 /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed by  
 Life Technologies. Contact : Feng Liang Life Technologies,  
 a division of Invitrogen 9600 Medical Center Drive  
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
 Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 174 a 286 c 315 g 180 t 13 others  
 ORIGIN

Query Match 38.8%; Score 828.6; DB 10; Length 968;  
 Best Local Similarity 94.4%; Pred. No. 1.4e-117;  
 Matches 923; Conservative 5; Mismatches 22; Indels 28; Gaps 6;

```

QY 1090 GCCACCCAGATGAGACCGCGGTACAGCCATCCCTTACCTACGAGCAGCTGAGCTCCTA 1149
D 954 GCCACCCAGATGAGACCGCGGTACAGCCATCCCTTACCTACGAGCAGCTGAGCTCCTA 896
QY 1150 AAGCATTAAGTGAATGAG-CTCTACCCACAGAGTTACCCGAGCTGATCCAGACAC- 1207
D 895 AAGCATTAAGTGAATGAGCTCTACCCACAGAGTTACCCGAGCTGATCCAGACACG 836
QY 1208 TGGCTACCTCTTCTCAAGATGAGCCCTGAGACATTCCGCAAGTGAATGTGACGTCC 1267
D 835 TGGCTACCTCTTCTCAAGATGAGCCCTGAGACATTCCGCAAGTGAATGTGACGTCC 776
QY 1268 TGGAGACCCCTG-AAAGCTTCTGTAAGTGCAGCAAGAGGACGAAATGATCTCAGGCT 1326
D 775 TGGAGACCCCTGAAAGGCTTCTGAAATGAAAGAGGACGAA----- 728
QY 1327 CCTGGGGGGCCCTCCACAGGTGGCCACCTGATGACGCGCTTGTGAAGGAGAGGGG 1386
D 729 -----ATGACTCTCAGAGTGGCCACCTGATGACGCGCTTGTGAAGGAGAGGGG 679
QY 1387 CAGCTAGACAAAGACACCTAGACACCTGACGCTTCTTACCTCTGGGTACTGTGCTCC 1446
D 678 CAGCTAGACAAAGACACCTAGACACCTGACGCGCTTCTTACCTCTGGGTACTGTGCTCC 619
QY 1447 CTGAGCCC-CGAGAGAGTGAAGTCTGTCGCCCGCAGAGAGATGAGGCGGTGAGGCCCA 1505
D 618 CTGAGCCCGAGAGAGTGAAGTCTGTCGCCCGCAGAGAGATGAGGCGGTGAGGCCCA 559
QY 1506 GAGCTGGAGACGCTGATGACCCCAAGGAGCTGAGAGCTCTATCCCAAGGCGCCCTTGC 1565
D 558 GAGCTGGAGACGCTGATGACCCCAAGGAGCTGAGAGCTCTATCCCAAGGCGCCCTTGC 499

```



```

QY 1566 TTTCAGACATGAAAGGCTCCGAATCTTGTGAGATCCAGTCTCTGAGTGGGC 1625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 498 TTTCAGACATGAAAGGCTCCGAATCTTGTGAGATCCAGTCTCTGAGTGGGC 439

QY 1626 CCCACGAGAGATTGAAGGCGCTCAGTACAGATGTGACATGTGACCTGGCCAGCTT 1685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438 CCCACGAGAGATTGAAGGCGCTCAGTACAGATGTGACATGTGACCTGGCCAGCTT 379

QY 1686 CATGAAGCTGGGAGAGATGGGCGTGGCCGCTGACCTGTGCTGAGTGCAGAACTCTT 1745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 378 CATGAAGCTGGGAGAGATGGGCGTGGCCGCTGACCTGTGCTGAGTGCAGAACTCTT 319

QY 1746 GGGACCCACGAGAGAGGCTGGAAGGCGAGAGAGCGGACCGCGGTGGCGACTGGAT 1805
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 318 GGGACCCACGAGAGAGGCTGGAAGGCGAGAGAGCGGACCGCGGTGGCGACTGGAT 259

QY 1806 CTTACGCGAGCGGAGAGACGACTGAGACAGCTGGGCGCTACAGGCGGAGATCCC 1865
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 258 CTTACGCGAGCGGAGAGACGACTGAGACAGCTGGGCGCTACAGGCGGAGATCCC 199

QY 1866 CAAGGCTACCTGCTCCAGACCTCAGCGTCCAGAGACCTCTCCGAGGAGCCCTGCT 1925
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 CAAGGCTACCTGCTCCAGACCTCAGCGTCCAGAGACCTCTCCGAGGAGCCCTGCT 139

QY 1926 CCTAGGACCTGAGACTGTTCTACCGTCTGACCTGCTTACCTCCAGCTGGCCTG 1985
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138 CCTAGGACCTGAGACTGTTCTACCGTCTGACCTGCTTACCTCCAGCTGGCCTG 79

QY 1986 AGGCGCCACTCTCTGCTGCGCCAGCCCTGCTGGGAGTCCCGCCTGGCCAGAGCAG 2045
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78 AGGCGCCACTCTCTGCTGCGCCAGCCCTGCTGGGAGTCCCGCCTGGCCAGAGCAG 19

QY 2046 GCACGGGTGATCCCCGTT 2063
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18 GCACGGGTGATCCCCGTT 1

RESULT 8
AL578289 953 bp mRNA ESI 16-FEB-2001
LOCUS AL578289 LTI_NFL006.Pl2 Homo sapiens cDNA clone CS0DK002YC06 3
DEFINITION AL578289 LTI_NFL006.Pl2 Homo sapiens cDNA clone CS0DK002YC06 3
ACCESSION AL578289
VERSION AL578289.1 GI:12942219
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 953)
AUTHORS L.L.W.B., Gruber,C., Jessée,J. and Polayès,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
    source
        1..953
            location/Qualifiers
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="CS0DK002YC06"
                /clone_11b="LTI_NFL006.Pl2"
                /tissue-type="placenta"
                /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                    was primed with a NotI-oligo(dT) primer. Five prime end
                    enriched, double-stranded cDNA was digested with Not I and
                    cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                    vector. Library was normalized; Feng Liang Life Technologies,
                    a division of Invitrogen 9800 Medical Center Drive
                    Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                    Email : fliang@lifetech.com URL :

```

```

BASE COUNT. 165 a 278 c 317 g 183 t 10 others
ORIGIN
Query Match 38.4%; Score 821.4; DB 10; Length 953;
Best Local Similarity 95.2%; Pred. No. 1,76-116;
Matches 919; Conservative 10; Mismatches 6; Indels 30; Gaps 7;
http://fulllength.invitrogen.com"

QY 1090 GCCACCCAGATGAGACCGCGTGAAGCCATCCCTTCACCTACGAGACCTGGAGCTCTTA 1149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 937 GCCACCCAGATGAGACCGCGTGAAGCCATCCCTTCACCTACGAGACCTGGAGCTCTTA 879

QY 1150 AAGATTAACCTGATGAGCTTACCCACAGATTACCCGATGATGATGATGATGATGAT 1209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 878 AAGATTAACCTGATGAGCTTACCCACAGATTACCCGATGATGATGATGATGATGAT 821

QY 1210 GGCTACCTCTTCTCAGATGAGCTGAGACATTCGAGTGGATGATGATGATGATGATGAT 1269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 822 GGCTACCTCTTCTCAGATGAGCTGAGACATTCGAGTGGATGATGATGATGATGATGAT 763

QY 1270 GAGACCTGGAAGCTTTCCTTGAAGTGACAAAGGCGACGAATGATCTCAGGCTCTT 1329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 762 KAGACCTGGAAGCTTTCCTTGAAGTGACAAAGGCGACGAATGATCTCAGGCTCTT 710

QY 1330 CGGGGGCCCCCACCAGAGTGGCCACCCTGATCGACCGCTTGTGAAGAGAGAGGGCCAG 1389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 711 -----CAGTGGCCACCCTGATCGACCGCTTGTGAAGAGAGAGGGCCAG 667

QY 1390 CTAGACAAAGACACCTAGACACCTGACDGCCTTCTACCTCGGTGATCTGCTCCTC 1449
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 666 CTAGACAAAGACACCTAGACACCTGACDGCCTTCTACCTCGGTGATCTGCTCCTC 607

QY 1450 AGCCCCGAGAGAGCTGAGTCCGTGCGCCCGCCAGCAGCATCTGGGGGATAGGCCCCAGGAC 1509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 606 AGCCCCGAGAGAGCTGAGTCCGTGCGCCCGCCAGCAGCATCTGGGGGATAGGCCCCAGGAC 547

QY 1510 CTGGACAGTGTGACCCCAAGGACAGTGGACCTGCTCTATCCCAAGGCGCGCTTGCCTTC 1569
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 546 CTGGACAGTGTGACCCCAAGGACAGTGGACCTGCTCTATCCCAAGGCGCGCTTGCCTTC 487

QY 1570 CAGAACATGAAGCGGGTCCGAATACTCTGTGAAGATCCAGTCTCTCTGGTGGGGCCCCC 1629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 486 CAGAACATGAAGCGGGTCCGAATACTCTGTGAAGATCCAGTCTCTCTGGTGGGGCCCCC 427

QY 1630 ACGGAGATTGGAAGCGGCTCAGTACGAGATGTGACATGATGATGATGATGATGATGAT 1689
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 426 ACGGAGATTGGAAGCGGCTCAGTACGAGATGTGACATGATGATGATGATGATGATGAT 367

QY 1690 AAGCTGCGAGAGATGCGGTGCTGCGCTGCTGACTGTGAGTGCAGAACTTCTGGGA 1749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 366 AAGCTGCGAGAGATGCGGTGCTGCGCTGCTGACTGTGAGTGCAGAACTTCTGGGA 307

QY 1750 CCCACGAGAGGCGCTGAAGGCGAGAGAGCGCACCGCGCGTGGGAGCTGATCTTA 1809
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 306 CCCACGAGAGGCGCTGAAGGCGAGAGAGCGCACCGCGCGTGGGAGCTGATCTTA 247

QY 1810 CGGAGAGCGGAGAGAGCTGAGACAGCTGGGGCTGGGGGCTACAGAGGGGAGATCCCAAC 1869
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 246 CGGAGAGCGGAGAGAGCTGAGACAGCTGGGGCTGGGGGCTACAGAGGGGAGATCCCAAC 187

QY 1870 GGCTACCTGCTCTAGACCTAGGCTGCAAGAGACCTCTGCGGGAGCCCTGCTCTCTTA 1929
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 186 GGCTACCTGCTCTAGACCTAGGCTGCAAGAGACCTCTGCGGGAGCCCTGCTCTCTTA 127

QY 1930 GGACCTGAGCTGTTCTACACCTGCTGAGCTGCTCTACCTCCACCCTGAGAGG 1989
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 GGACCTGAGCTGTTCTACACCTGCTGAGCTGCTCTACCTCCACCCTGAGAGG 67

QY 1990 CCCCACTCCCTGCTGGCGCCAGCCCTGCTGAGATGCCCTGGCCAGAGAGAGGAC 2049
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 CCCCACTCCCTGCTGGCGCCAGCCCTGCTGAGATGCCCTGGCCAGAGAGAGGAC 8

```

QY	2050	GGGTG	2054
Db	7	GGGTG	3
RESULT	9		
AL556862			
LOCUS			
DEFINITION	AL556862 LTL_NFL006_PL2 Homo sapiens cdna clone CS0DK012YE02 5	EST	16-FEB-2001
ACCESSION	AL556862		
VERSION	AL556862.1	GI:12899917	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191, 91006 EVRY cedex - France		
	Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES			
Source	1..965		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="CS0DK012YE02"		
	/clone_lib="LTL_NFL006_PL2"		
	/issue_type="Placenta"		
	/note="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA		
	was primed with a NotI-Oligo(dT) primer. Five prime end		
	enriched, double-stranded cDNA was digested with Not I and		
	cloned into the Not I and Eco RV sites of the pCMVSPORT 6		
	vector. Library was normalized. Library was constructed by		
	Life Technologies. Contact : Feng Liang Life Technologies,		
	a division of Invitrogen 9800 Medical Center Drive		
	Rockville, Maryland 20850, USA Fax : (1) 301 610 8371		
	Email : fliang@lifetech.com URL :		
	http://fulllength.invitrogen.com"		
BASE COUNT	135 a 340 c 314 g 171 t 5 others		
ORIGIN			
Query Match	38.1%; Score 815.4; DB 10; Length 965;		
Best Local Similarity	95.5%; Pred. No. 1.4e-115;		
Matches	920; Conservative 5; Mismatches 26; Indels 12; Gaps 8;		
QY	33	TGTGACGCGCGGACAGAGACTACCGGTGAGCCACGCTCCCTCGGTGATCTAC	92
Db	1	TGTGACGCGCGGACAGAGACTACCGGTGAGCCACGCTCCCTCGGTGATCTAC	60
QY	93	ACAGACCATGGGCTTG-CAAGGGCTCGACCCCTGTTGCTCTGTGGGACCGCC- ----	146
Db	61	ACAGACCATGGGCTTGCCAAAGGCTCGACCCCTGTTGGGGCTCTGTGGGACCGCCCT	120
QY	147	TGGCAGCCTCGTTCGCTTGTAGCCTGGATGGGGGACATCCCGGAGAGCCCTGGC	206
Db	121	CGGACGCTCTCTGTTCTCTCTTAAAGCTCGATGGGGGAGCCCTCGAGGACCTGGC	180
QY	207	TGGAGAGACAGGAGAGAGTGGCCCGCTGGGGGAGAGCTCGACACCCCATTAACAT	266
Db	181	TGGAGAGACAGGAGAGAGAGTGGCCCGCTGGAGAGAGTGGCCCAACCCATTAACAT	239
QY	267	TTTCAGGCTCTCCCTCGGCAACTCTTGGCTTCCCGTGTGGAGAGTGTCCGGCTGAG	326
Db	240	TTTCAGGCTCTCCCTCGGCAACTCTTGGCTTCCCGTGTGGAGAGTGTCCGGCTGAG	299
QY	327	CACGAGGCTGTCCGGGAGCTGGCTGTGGCTTGGACAGAAAGATGTCAACCTCTCAAC	386
Db	300	CACGAGGCTGTCCGGGAGCTGGCTGTGGCTTGGACAGAAAGATGTCAACCTCTCAAC	359

QY	387	AGAGAGCGTGGCGCTCTGAGCGACACGGCTCTCTGAGCGCCCGCAGGACGCG	-GACGCGC	445
Db	360	AGAGAGCTGGCGCTCTGAGCGACACGGCTCTCTGAGCGCCCGCAGGACGCG	CGC	419
QY	446	TCGCATTGAGACCTGCTGCTATTTCCTCAACCCAGATGCTTCTCGAGGCGCCAGGCTGCA	505	
Db	420	TCGCATTGAGACCTGCTGCTATTTCCTCAACCCAGATGCTTCTCGAGGCGCCAGGCTGCA	479	
QY	506	CCCGTTTCTTCTCCCGCATCACGAAGGCCATGTGACCTCTCTCCGAGGGGGCTCCCG	565	
Db	480	CCCGTTTCTTCTCCCGCATCACGAAGGCCAATGTGACCTCTCTCCGAGGGGGCTCCCG	539	
QY	566	ACGCACAGGGGCTGCTGCTGGGCGCTGAGCGCTGAGCGTGGGGGTGTGGGGGCTCTGCTGA	625	
Db	540	ACGCACASGGGCTGCTGCTGGGCGCTGAGCGCTGAGCGTGGGGGTGTGGGGGCTCTGCTGA	599	
QY	626	GGAGAGCTGATGTGGGGGCTGTGGAGGCGCTGAGCTTGGCGACCTGCTGAGCGCTTTGTGG	685	
Db	600	GGAGAGCTGATGTGGGGGCTGTGGAGGCGCTGAGCTTGGCGACCTGCTGAGCGCTTTGTGG	659	
QY	686	CCGAGTCGGCGCACTGCTGCTACCCCGCTGTGTGACCTGCTCCCGGGACCCCTGGACCAAG	745	
Db	660	CCGAGTCGGCGCACTGCTGCTACCCCGCTGTGTGAACTGCTCCCGGACCCCTGGACCAAG	719	
QY	746	ACCACAGAGGACACCAAGGGGGGCTGTGACAGGGGGGGGAGCCCCCTAGCGGCCCGCGT	805	
Db	720	ACCACAGAGGACACCAAGGGGGGCTGTGACAGGGGGGGGAGCCCCCTAGCGGCCCGCGT	778	
QY	806	CGACATGCTGTCTCCACGATGACGCTCTGCGGGGCGCTGCTCCGCTGCTGAGGCCAGC	865	
Db	779	CGACATGCTGTCTCCACGATGACGCTCTGSGGGGCGCTGCTCCGCTGCTGAGGCCAGC	838	
QY	866	CCATATATCCGACGATCCCCGAGGGCATCTGTGCGCGCTGGCGGGCAACGCTCTCTGGGG	925	
Db	839	CCATATATCCGACGATCCCCGAGGATCCCGCA-GSACATCTGTGGCGGCTGTG-SGCAACGCTCTCTGGGG	896	
QY	926	ACCCATCTCTGGGGGAGGCTCAACGAGACCACTCTCCGCGCGGTTCCGCGGGAGAACTGG	985	
Db	897	ACCCATCTCTGGGGGAGGCTCAACGAGACCACTCTCCGCGCGGTTCCGCGGGAGAACTGG	955	
QY	986	AGA 988		
Db	956	AGA 958		
RESULT 10				
LOCUS	ALS14070	912 bp	mRNA	EST
DEFINITION	ALS14070 LTI_NFL006.PL2	Homo sapiens cDNA clone Cl0BA004ZG08.5		
ACCESSION	ALS14070	prime, mRNA sequence.		
VERSION	ALS14070.1	GI:12777564		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	Li,W.B., Gruber,C., Jesssee,J. and Polayes,D.			
JOURNAL	Full-length cDNA libraries and normalization			
COMMENT	Unpublished (2001)			
FEATURES	Contact: Genoscope			
source	Genoscope - Centre National de Sequencage			
	BP 191 91006 Evry cedex - France			
	Email: sequele.genoscope.cns.fr, web : www.genoscope.cns.fr.			
	Location/Qualifiers			
	1..912			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="Cl0BA004ZG08"			
	/clone_id="LTI_NFL006.PL2"			
	/tissue_type="placenta"			



```

|||||
Db 606 CCAAGCAGCTGAGCTCTCTATCCCAAGGCCCGCTTGTCCAGAACATGAACGGG 547
Qy 1585 TCCGAATCTTCGTGAATATCCAGTCTTCTGGGTGGGGCCCCCAGAGAGATTGAAG 1644
Db 546 TCCGATATCTTCGTGAATATCCAGTCTTCTGGGTGGGGCCCCCAGAGAGATTGAAG 487
Qy 1645 GCGCTCATCAGCAGAAATGTAGCATGTGGCTGGCCAGTTCATGAAGCTGGCAGCAT 1704
Db 486 GCGCTCATCAGCAGAAATGTAGCATGTGGCTGGCCAGTTCATGAAGCTGGCAGCAT 427
Qy 1705 GCGGTCTCCCTGTCAGTCTGTGGCTGAGTGCAAGAACTTGTGGAGCCCGACGTGGAGGC 1764
Db 426 GCGGTCTCCCTGTCAGTCTGTGGCTGAGTGCAAGAACTTGTGGAGCCCGACGTGGAGGC 367
Qy 1765 CTGAAGGGGAGAGAGCGCCCGCCCGGGTGGGAGCTGGATCTTCAAGCGGAGAGAC 1824
Db 366 CTGAAGGGGAGAGAGCGCCCGCCCGGGTGGGAGCTGGATCTTCAAGCGGAGAGAC 307
Qy 1825 GACCTGACAGCGTGGGCTGGGCTGACAGGGCGGATCCCAACGCGTACCTGGCTCA 1884
Db 306 GACCTGACAGCGTGGGCTGGGCTGACAGGGCGGATCCCAACGCGTACCTGGCTCA 247
Qy 1885 GACCTGACAGCGTGGGCTGGGCTGACAGGGCGGATCCCAACGCGTACCTGGCTCA 1944
Db 246 GACCTGACAGCGTGGGCTGGGCTGACAGGGCGGATCCCAACGCGTACCTGGCTCA 187
Qy 1945 CTCACCGTCTGGGCTGCTCTCTACCTGCAACCGCTGAGGGCGCCACCTCTCTCT 2004
Db 186 CTCACCGTCTGGGCTGCTCTCTACCTGCAACCGCTGAGGGCGCCACCTCTCTCT 127
Qy 2005 GCGCCCAAGCCCTGCTGGGATCCCGCTGCGCAGAGCAGGAGGAGTCCCTTC 2064
Db 126 GCGCCCAAGCCCTGCTGGGATCCCGCTGCGCAGAGCAGGAGGAGTCCCTTC 67
Qy 2065 CACCCCAAGAGAACTCGGCTGCTGATTAACGGGAGATGCCCCCTGACAGACAAAAA 2124
Db 66 CACCCCAAGAGAACTCGGCTGCTGATTAACGGGAGATGCCCCCTGACAGACAAAAA 7
Qy 2125 AAAAAA 2130
Db 6 AAAAAA 1

RESULT 12
AL555912 817 bp mRNA EST 16-FEB-2001
LOCUS AL555912 LTI.NFL006.PL2 Homo sapiens cDNA clone CS0DK002YC06 5
DEFINITION AL555912 LTI.NFL006.PL2 Homo sapiens cDNA clone CS0DK002YC06 5
ACCESSION AL555912
VERSION AL555912.1 GI:12898034
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 817)
AUTHORS L.I.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
Location/Qualifiers
1..817
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002YC06"
/clone_1db="LTI.NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

```

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : [filang@lifetech.com](mailto:filang@lifetech.com) URL : <http://fulllength.invitrogen.com>

BASE COUNT 109 a 285 c 270 g 151 t 2 others  
ORIGIN

Query Match 31.7%; Score 677.8; DB 10; Length 817;  
Best Local Similarity 95.0%; Pred. No. 1.6e-94;  
Matches 765; Conservative 1; Mismatches 28; Indels 11; Gaps 6;

```

Qy 46 CAGAGAGCTACCGGTGAGACCCAGCGTCCCTCCCTGGGATCTACACAGACATGSCC 105
Db 1 CAGAGAGCTACCGGTGAGACCCAGCGTCCCTCCCTGGGATCTACACAGACATGSCC 60
Qy 106 TTG-CAAGGGCTCGACCCCTGTTGCTCTGTGGGAGCGGCC-----TGGCAGCTCTG 159
Db 61 TTGCAAGGGCTCGACCCCTGTTGCTCTGTGGGAGCGGCCCGCCCTGCAAGCTCTCTG 120
Qy 160 TTCTGCTCTTCAAGCTCGATGGTGCATCCCGCAGAGACCTGGCTGAGAGACAGAG 219
Db 121 TTCTGCTCTTCAAGCTCGATGGTGCATCCCGCAGAGACCTGGCTGAGAGACAGAG 180
Qy 220 ACGAGTCTGCCCCCTGGGGGAGTCTGACACACCCCATATACATTTCCAGCTCTTC 279
Db 181 CAGGAGCTGCGCCCTCGAGGAGTCTGGCCACCCAC-TCACATTTCCAGCTCTTC 239
Qy 280 CCTGCGCAACCTCTGGGCTCCCGTGTGGAGAGTCTCGGCGTGGAGACGAGAGTTC 339
Db 240 CCTGCGCAACCTCTGGGCTCCCGTGTGGAGAGTCTCGGCGTGGAGACGAGAGTTC 299
Qy 340 CGGAGCTGCTGTGGCCCTTGGCAGAGAAATGTCAAGCTCTCAACAGACAGCTGGC 399
Db 300 CGGAGCTGCTGTGGCCCTTGGCAGAGAAATGTCAAGCTCTCAACAGACAGCTGGC 359
Qy 400 TGTCTGGCTCACCGGCTCTGTAGCCCCCGAGACCTGGAGCCCTCCATTGGACCTG 459
Db 360 TGTCTGGCTCACCGGCTCTGTAGCCCCCGAGACCTGGAGCCCTCCATTGGACCTG 419
Qy 460 CTGCTATTCTCAACCCAGATGCGTTTCGCGGCCCGCCAGGCTGACCCGTTCTTC 519
Db 420 CTGCTATTCTCAACCCAGATGCGTTTCGCGGCCCGCCAGGCTGACCCGTTCTTC 479
Qy 520 CGCATCACGAAGCCCAATGTGAGCTGCTCCGAGGGGGGCTCCCGAGAGCAGAGCG-CT 578
Db 480 CGCATCACGAAGCCCAATGTGAGCTGCTCCGAGGGGGGCTCCCGAGAGCAGAGCGNCT 539
Qy 579 GCTGCTCGGCTCTGCGCTGCTGGGGTGTGGGGGCTCTGCTGAGCAGAGCTGATGT 638
Db 540 GCTGCTCGGCTCTGCGCTGCTGGGGTGTGGGGGCTCTGCTGAGCAGAGCTGATGT 599
Qy 639 GCGGCTCTGGGAGGCGTGGCTGGAGCTGCTGGGGGCTTTGGGCGAGTGGGCGGA 698
Db 600 GCGGCTCTGGGAGGCGTGGCTGGAGCTGCTGGGGGCTTTGGGCGAGTGGGCGGA 659
Qy 699 AGTGTGCTACCCCGGCTGTGAGTGGCCCGGAGCCCTTGAGCAGAGCAGCAGAGAGC 758
Db 660 AGTGTGCTACCCCGGCTGTGAGTGGCCCGGAGCCCTTGAGCAGAGCAGCAGAGAGC 719
Qy 759 AGCCAGGGCGGCTGTGAGGGCGGGGAGCCCTTACGGCCCCCGTGTGACATGCTGT 818
Db 720 AGCCAGGGCGGCTGTGAGGGCGGGGAGCCCTTACGGCCCCCGTGTGACATGCTGT 777
Qy 819 CTCACGATGAGAGCTCTGCGGGGC 843
Db 778 CT-CACGATGAGAGCTCTGCGGGGC 801

```

	RESULT	13
LOCUS	A1813749/c	
DEFINITION	A1813749 mRNA EST 07-MAR-2000 WR79605.x1 NCI_CGAP_PanI Homo sapiens cDNA clone IMAGE:2421584 3'	
ACCESSION	Similar to TR:Q14859 Q14859 PRE-PRO-MEGAKARYOCYTE POTENTIATING- FACTOR PRECURSOR.; , mRNA sequence.	
VERSION	A1813749	
KEYWORDS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo. A1813749 GI:5424964	
SOURCE	Homo sapiens human.	
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Life Technologies catalog #: 11548-013 DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP Clone Distribution Information can be found through the I.M.A.G.E. Consortium/LML at: <a href="#">www.bio.lml.gov/bnrf/image/image.html</a> Insert Length: 2146 Std Error: 0.00 Seq primer: -40bp from Glpco High quality sequence stop: 424. Location/Qualifiers	
FEATURES	source 1..737 organism="Homo sapiens" db_xref="taxon:9606" clone_image="IMAGE:2421584" clone_id="NCI.CGAP.PanI" tissue_type="adenocarcinoma" lab_host="DHIOB" note="Organ: pancreas; Vector: pCMV-Sport6; Site_1: SalI, Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT." Average Insert size 1.72 kb. Life Technologies catalog #: 11548-013"	
BASE COUNT	126 A    230 C    241 G    139 T    1 others	
ORIGIN		
Query Match	31.7%; Score 677; DB 10; Length 737;	
Best Local Similarity	95.7%; Pred. No. 2.2e-94;	
Matches 706; Conservative	0; Mismatches 31; Indels 1; Gaps 1;	
OY	1377 GGGAAGGGGCCAGCTAGACAAGAACCCTTAGACACTGTACCGCCTTCCTACCTGGGTA	1436
Dd		
OY	737 GGGCCGGGGCAAGCTAGAAAAGAACCTTAACTCCCTTTATCTGGGGTA	679
OY	1437 CCTGCATCCCACACCCCCGAGAGACTAGAGCTCGTGCCGCCGCCCACAGCATCTGGGGGT	1496
Dd		
OY	678 CTGTGCTCCTCAAGCCACGAGAACTTAGCTCCGTGCCCTAGCTAGCATCTGGGGGT	619
OY	1497 CAGGCCCCAGGACCTGACACAGCTGTGACCCAAGGACGATGAGCTCTATTCCAAAGCC	1556
Dd		
OY	618 CAGGCCCCAGGACCTGACACAGCTGTGACCCAAGGACGATGAGCTCTATTCCAAAGCC	559
OY	1557 CCAGCTTCCTTCCAGAACATGAAACGGGTCGGAATCTTCTGTAAGATCACTCTTCT	1616
Dd		
OY	558 CCAGCTTCCTTCCAGAACATGAAACGGGTCGGAATCTTCTGTAAGATCACTCTTCT	499
OY	1617 GGGTGGGGCCCCACAGGAGGAGATTGAAGCGCTAGCTAGCAAGANTGAGATGGAGCT	1676
Dd		
OY	498 GGGTGGGGCCCCACAGGAGGAGATTGAAGCGCTAGCTAGCAAGANTGAGATGGAGCT	439
OY	1677 GGCCACTTTCATGAGAGCTGCGGACGATGCGCTGCTGCTTGAAGATGAGAGTGA	1736
Dd		
OY	438 GGCCACTTTCATGAGAGCTGCGGACGATGCGCTGCTGCTTGAAGATGAGAGTGA	379
OY	1737 GAACCTTTCGAGACCCCAAGTGAAGGCTGAGAGCGGAGGACGCGCAGCGCGGATGCG	1796

Db	378	GAACCTTCTGGGACCCCAACGCTGGAGGGCTTAAGCGGAGGAGCGGACCGCCCGGCTGC	319
QY	1797	GGACTGATCTCTACGGCAGCGGCAGAGACGACCTTGACACGCTGGGGCTGGGGCTACACGG	1856
Db	318	GGACTGATCTCTACGGCAGCGGCAGGAGGACGACTGTGGACACGCTGGGGCTGGGGCTACACGG	259
QY	1857	CGGCATCCCCCAAGCGCTACCTGGTCTCTAGACCTCAGCGCTGCAAGAGACCTCTGGGGAC	1916
Db	258	CGGCATCCCCCAAGCGCTACCTGGTCTCTAGACCTCAGCATGCAAGAGGCTCTGGGGAC	199
QY	1917	GGCCGTGCTCTAGACCTGGACCTGTTCTCACCCTCTGGACACGCTCAGCTCAGCTCCAC	1976
Db	198	GGCCGTGCTCTAGACCTGGACCTGTTCTCACCCTCTGGACACGCTCAGCTCAGCTCCAC	139
QY	1977	CGTGGCTTGAGGGGCCCCCACTCCCTTGCTGAGGCCAGCCCTCTGGGGATCCCGCTGGC	2036
Db	138	CGTGGCTTGAGGGGCCCCCACTCCCTTGCTGAGGCCAGCCCTCTGGGGATCCCGCTGGC	79
QY	2037	CAGAGCAGGAGCAGGGGTATCCCGCTTCACCCCAAGAGAACTGGCGCTCAGTAACGGG	2096
Db	78	CAGAGCAGGAGCAGGGGTATCCCGCTTCACCCCAAGAGAACTGGCGCTCAGTAACGGG	19
QY	2097	AACATGCCCCCTGGCAGAC	2114
Db	18	AACATGCCCCCTGGCAGAC	1
RESULT	14		
LOCUS	AI984029/c		
DEFINITION	wf52e09.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:251112 3'	EST	09-MAR-2000
DESCRIPTION	similar to TR:Q14859 Q14859 PRE-PRO-MEGAKARYOCYTE POTENTIATING FACTOR PRCURSOR. ; contains MSRI.t2 MSRI repetitive element ; , mRNA sequence.		
ACCESSION	AI984029		
VERSION	AI984029.1	GI:5811248	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 759)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997)		
FEATURES	Source		
FEATURES	Location/Qualifiers		
SOURCE	1..759		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_image="IMAGE:251112"		
	/clone_id="NCI_CGAP_Pan1"		
	/tissue_type="adenocarcinoma"		
	/lab_host="DH10B"		
	/note="Organ: pancreas; Vector: pCMV-Sport6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.72 kb. Life Technologies catalog #:		
	11548-013"		
BASE COUNT	131 a	237 c	232 g 155 t 4 others
ORIGIN			

Query Match 30.4%; Score:650.2; DB 10; Length 799;  
 Best Local Similarity 93.4%; Pred. No. 2.7e-90;  
 Matches 709; Conservative 0; Mismatches 47; Indels 3; Gaps 3;

QY 1366 CGCTTTGGAAGGGAAGGGCCAGCTAGACAAAGACACCTTAGACACCTGACCCCTTC 1425  
 DB 757 CGCTTTTGGAAAGGAAGGGCCAGGTAGAC-AMGACACCTAAGACACCTTGGCCCTTTN 699  
 QY 1426 TACCTGGAGTACCTGCTCCCTCAGCCCCAGAGTGTAGCTGCTCCGCCCGGAGAGC 1485  
 DB 698 TA-CTGGGTACTGCTGCTCCTCAGGCCCGGAGAGAGTGAAGTCCGGTCCCGGAGAGC 640  
 QY 1486 ATCTGGGCGGTAGAGCCCGGAGAGCTGAGACAGTGTGACCCAGGAGAGCTGAGCTCTC 1545  
 DB 639 ATTGGGGGGGTAGAGCCCGGAGAGTGTGACAGTGTGACAGGAGTGTGAGAGCTGCTTC 580  
 QY 1546 TATCCCAAGGCGCCCTTGTCTTCCAGAACATGAACGGGTCCGAATCTTGTGTGAAGTTC 1605  
 DB 579 TATCCCAAGGCGCCCTTGTCTTCCAGAACATGAACGGGTCCGAATCTTGTGTGAAGTTC 520  
 QY 1606 CAGTCTTCTGAGTGGGGGCCCCGAGAGATTGAAGGGGCTCAGTACAGAGATGTG 1665  
 DB 519 CAGTCTTCTGAGTGGGGGCCCCGAGAGATTGAAGGGGCTCAGTACAGAGATGTG 460  
 QY 1666 AGCATGACCTTGGCCAGCTTCAATGAAGCTGCGAGAGATGCGGTGCTCCGTTGACTGTG 1725  
 DB 459 AGCATGACCTTGGCCAGCTTCAATGAAGCTGCGAGAGATGCGGTGCTCCGTTGACTGTG 400  
 QY 1726 GCTGAGGAGCAAACTCTGAGGAGCCGAGTGGAGGCGCTGAAGGGCGAGAGGCGGAC 1785  
 DB 399 GCTGAGGAGCAAACTCTGAGGAGCCGAGTGGAGGCGCTGAAGGGCGAGAGGCGGAC 340  
 QY 1786 CGCCCGGTGCGGAGTGTGATCTTACGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1845  
 DB 339 CGCCCGGTGCGGAGTGTGATCTTACGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 280  
 QY 1846 GGGCTACAGGGCGGATCCCAAGAGGCTACTGCTCTAGACCTGAGCTGAGAGAGAGAG 1905  
 DB 279 GGGCTACAGGGCGGATCCCAAGAGGCTACTGCTCTAGACCTGAGCTGAGAGAGAGAG 220  
 QY 1906 CTCTC-GGAGAGCGGCTCTCTCTAGAGAGTGTGAGTGTCTGAGAGAGAGAGAGAG 1964  
 DB 219 CTCTC-GGAGAGCGGCTCTCTCTAGAGAGTGTGAGTGTCTGAGAGAGAGAGAGAG 160  
 QY 1965 CTAAGCTTCAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2024  
 DB 159 CTAAGCTTCAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 100  
 QY 2025 TCCCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2084  
 DB 99 TCCCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 40  
 QY 2085 TCAGTAACGGGAGACATGCCCCCTGAGACAAAAAAA 2123  
 DB 39 TCAGTAACGGGAGACATGCCCCCTGAGACAAAAAAA 1

RESULT 15  
 A1744747/c 799 bp mRNA EST 21-JUN-1999  
 LOCUS t114b11.x1 NCI-CGAP\_Ov23 Homo sapiens cDNA clone IMAGE:2218269 3'  
 DEFINITION s1milar to TR:Q14859 Q14859 PRE-PRO-MEAKARYOCYTE POTENTIATING  
 FACTOR PRECURSOR; contains MSRL.t2 MSRL repetitive element; mRNA  
 sequence.  
 ACCESSION A1744747  
 VERSION A1744747.1 GI:5113035  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 799)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LML at:  
[www-bio.lml.nih.gov/dbirp/image/image.html](http://www-bio.lml.nih.gov/dbirp/image/image.html)  
 Seq primer: -400P from Glibco  
 High quality sequence stop: 403.  
 Location/Qualifiers  
 1..799  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2218269"  
 /issue\_type="tumor, 5 pooled (see description)"  
 /lab\_host="DH10B"  
 /note="Organ: ovary; Vector: pCMV-Sport6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.35 kb. Tumor types include: mixed  
 Mullerian tumor, papillary serous, clear cell, spindle  
 cell. All are primary tumors, metastasis positive. Life  
 Technologies catalog #: 11534-013"

BASE COUNT 133 a 244 c 266 g 150 t 6 others  
 ORIGIN

Query Match 30.1%; Score 643.8; DB 10; Length 799;  
 Best Local Similarity 94.0%; Pred. No. 2.5e-89;  
 Matches 721; Conservative 0; Mismatches 41; Indels 5; Gaps 5;

QY 1351 GCCACCTGATGACCGCTTTGTGAAGGAAGGGCCAGTACAGAAAGACACCTAGAC 1410  
 DB 767 GCCCNCCTGATGCGCCGCTTTGTAAAGGAAGGCAAGGAGTACAGACACCTAGAGACA 708  
 QY 1411 ACCGACACCGCTT-CTACCTGAGTACCTGTG-CTCCCTAGGCCCGG-AGAGAGTGAGC 1467  
 DB 707 ACCTGACCCGCTTTCCTGAGGAGTGTGCTGCTCCTCAGGACGAGAGAGTGTGAGC 648  
 QY 1468 TCCGAGCCCGGAGAGATGCGGCGGTGAGGCGCCGAGAGAGTGTGAGAGCA 1527  
 DB 647 TCCGAGCCCGGAGAGATGCGGCGGTGAGGCGCCGAGAGAGTGTGAGAGCA 588  
 QY 1528 AGGAGCTGAGAGTCTCTATCCCAAGGCCCGCTTGC-TTTCAGAAATGAAGAGGCTC 1586  
 DB 587 GGGAGCTGAGAGTCTCTATCCCAAGGCCCGCTTGC-TTTCAGAAATGAAGAGGCTC 528  
 QY 1587 CGAATACTTGTGAAGATCCAGTCTTCTGAGTGTGAGGCGCCCGAGAGAGATTGAAGGC 1646  
 DB 527 CGAATACTTGTGAAGATCCAGTCTTCTGAGTGTGAGGCGCCCGAGAGAGATTGAAGGC 468  
 QY 1647 GCTCAGTACAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1706  
 DB 467 GCTCAGTACAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 408  
 QY 1707 GGTGCTGCGCTTGAAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 1766  
 DB 407 GGTGCTGCGCTTGAAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 348  
 QY 1767 GAAGGCGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1826  
 DB 347 GAAGGCGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 288  
 QY 1827 CCTGAGACAGCTGGGGGCTGAGGAGTACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1886  
 DB 287 CCTGAGACAGCTGGGGGCTGAGGAGTACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 228



**THIS PAGE BLANK (uspto)**





PI	Chang K, Pastan I;
DR	WPI; 1997-3/2620/34.
XX	DR N-PSDB; AAT91079.
XX	
XX	Agents for targeting mesothelin, a tumour cell antigen - used for
PT	the detection or inhibition of growth of e.g. mesotheliomas, ovarian
PT	cancers and squamous cell carcinomas
XX	
PS	Claim 1; Page 58-60; 72pp; English.
XX	
CC	This protein comprises the human CAK1 antigen which is found
CC	on mesothelium, mesotheliomas, ovarian cancers and some
CC	squamous cell carcinomas. The antigen has been designated
CC	mesothelin. Its amino acid sequence was deduced from a cDNA
CC	clone (see AAT91079) isolated from a HeLa cDNA library. A 40 kDa
CC	form (K1) of mesothelin that is found on the surface of human
CC	ovarian tumour OVCAR-3 cells appears to be derived from the 69
CC	kDa precursor by several processing steps. A claimed method for
CC	specifically delivering an effector molecule to a tumour cell
CC	bearing an antigen comprising at least 10 contiguous amino acids
CC	of mesothelin involves: (a) providing a chimeric molecule comprising
CC	the effector molecule attached to a targeting molecule that
CC	specifically binds to mesothelin; and (b) contacting the tumour with
CC	the chimeric molecule such that the chimeric molecule specifically
CC	binds to a tumour cell. Also claimed is a method for inhibiting
CC	mesothelin expression or activity by contacting mesothelin bearing
CC	cells with inhibitory nucleic acids for the mesothelin gene. The
CC	methods can be used to detect tumour cells and to inhibit the
CC	growth of cells bearing mesothelin. Mesothelin-derived antigens
CC	may be used in vaccines for the inhibition or prevention of
CC	mesotheliomas or ovarian tumours.
XX	
XX	Sequence 628 AA;

Query Match	Similarity	100.0%	Score 3261	DB 18	Length 628
Best Local	Similarity	100.0%	Pred. NO. 2.3e-289		
Matches 628	Conservative	0	Mismatches	0	Indels
					Gaps 0
QY 1	MALORLDPWCSCGDRPGSILFLFSLFSGVWHPARTLAGETGTSAPRLGCVLTTPHNISSLS	60			
DB 1	malqrlidpcwscgdrpgsllflfslfsgvwhparrtagetgtsaprgylttrphnissls	60			
QY 61	PROLLGPPCAEYSGCISTERVRELAVALAOKNVKLSSTEOCLRLAHRLSEPPEDLALPLD	120			
DB 61	prqligfpcaeysgcistervrelavalaknvklssteqrlclahrseppedaalpdl	120			
QY 121	LLFLNPDAFSGPOACTREFSRITKANVOLLPRGAEERQLRLPALACWGRVSLSEADY	180			
DB 121	llflnpdafsgpqactrfrstlckanvdlprgaerqrlrpalacwgrvslsleady	180			
QY 181	RALGGLACDLGREGFAESAEEVLLPRLVYCPGLDDQOEAAARALOGGPRVGPSTMSV	240			
DB 181	ralgglacdlpgrfrfaesaeavlprlvscpprlldqdgaeaaalqggpprvppstmsv	240			
QY 241	STMDALRGLRPLVLCOPILIRSIPOGIVAAWRORSSNDPSMRPERTILPFRFRREYEKTAC	300			
DB 241	stmdalrgrlrvlygprllrsipgglvaawqrssrdpswrgperrtilprfrfrevktaac	300			
QY 301	PSGKARREIDESLFYKKWELEACYDAALATOMDRVNAIPRTYQOVLVKNKIDELYPQ	360			
DB 301	psgkareideslffykkweleacvdaaalatmdrvnaiprtypqldvllknkldelypq	360			
QY 361	GYPSVIOHLELYLFKMSPERDIRKKNWVSLFTIKALLVDNGHENSQARRRRLQVATL	420			
DB 361	gypsvldhlyllflkmsperdrkknwvslftlkallevdghensqparrlpqvatl	420			
QY 421	IDREFKGQGLDKDLTDLTLTAFFPYGLCSLSEELSSVPPSSITAVWRRODLDTCOPROD	480			
DB 421	ldrfkvgqgltdktdldtcltalfpyglcslseelssvppssitavwrpddldtcprqld	480			
QY 481	VLYPKARLAFONMGSEYFVKIQSFLGCAPTEDLKALSOQNVSKDLATFMKLRTDAVPL	540			

[illegible]

RESULT	2
AA	AA53992
ID	AA53992 standard; Protein; 622 AA

AC AAR53992;

DT 08-DEC-1994 (first entry)

DE Megakaryocyte potentiator.

KM Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;

XX

[illegible]

Protein

Protein FT

XX  
XX  
E000410313-1

XX 11-MAY-1964  
DD

XX  
DE 25-OCT-1993.

XX  
DB 23-OCT-1992.

PR 11-NOV-1992;  
 DP 09-DEC-1992:

XX  
DA  
(CHHS ) CHING

XX	Hattori K K
XX	
PT	

XX WPT. 1994-167  
DB

DR N-PSDB; AAQ63  
XX

PT New mega:kary  
PT thrombocyte

PS Disclosure: F

XX DNA encoding

CC use in treatm  
XX

50	Sequence	62
----	----------	----

Query Match

Best Local Similarity: 601: C  
Matches

OV 1 MAJOR DE

Ph 1 maj ptarr

59 J.SPROLIGOV

61 1sprq11a

Sequence 622 AA;

Query Match 93.6%; Score 3053; DB 15; Length 622;

Best Local Similarity 55.4%; Freq. NO. 2.4e-270;  
Matches 601; Conservative 3; Mismatches 16; Indels 10; Gaps 3;

QY 1 MALQRLDP-CWSCGDRP-GSLLELLFSLGWHPARTLAGETGTESAPLGCVLTPHNISS 58

Db 1 malptarpllgscgtpalgslflflfslgwvpssrtlaetgqaaplidy lanpniss 60

QY 59 LSPRQLGFPCAEVSGLSTERVREIAVALAOKNVKLSTEQLRCLAHRLSEPPEDDALPL 11

Db 61 1sprql1gfpcaevsg1stervrelavalagknvk1steg1rc1ahr1sepedida1p1 12



DR WPI: 2000-572118/53.  
DR N-PSDB: AAA64366.  
PT Diagnosing carcinomas, e.g. adenocarcinoma or ovarian carcinomas, by  
PT detecting reactivity of a molecule (with an antigenic determinant and  
PT present in a sample) with an antibody specific for a mesothelin related  
PT antigen polypeptide  
XX  
PS Claim 48; Fig 7A-B; 92pp; English.  
XX  
CC The present sequence represents a human soluble mesothelin related (SMR)  
CC antigen. Antibodies specific to mesothelin related antigen (MRA)  
CC polypeptides are used for the detection of a malignant condition.  
CC Mesothelin is a differentiation antigen which is expressed in the  
CC surfaces of normal mesothelial cells and also on certain cancer cells,  
CC including epithelial ovarian tumours and mesotheliomas. The method is  
CC useful for detecting the presence of a malignant condition, specifically,  
CC adenocarcinoma, mesothelioma, ovarian carcinoma, pancreatic carcinoma  
CC or non-small cell lung carcinoma.  
XX  
SQ Sequence 402 AA;  
  
Query Match 47.4%; Score 1547; DB 21; Length 402;  
Best Local Similarity 97.1%; Pred. No. 7.5e-133;  
Matches 301; Conservative 1; Mismatches 0; Indels 8; Gaps 1;  
  
OY 291 FRREKTCACPSGKKAREIDSLFFKKWELEACVDALATQMDRVNAIPFTYEQDLYL 350  
DB 1 frreketacpsgkkaareideslffkkweleacvdaallatqmdrvnaipfyeqldvll 60  
OY 351 KHKIDELYPGYPSVSIQHLGYLFLKMSPEDIRKKNWTSLETLKALLEVDKGHEMSPOAP 410  
DB 61 kkkldelypgypsvsiqhlgyllfklmspedirknwtsletlkallevnkghems---- 116  
OY 411 RRPPLQVATLIDRFVKGKGGQLDKDLDLTLTAFYPGYLCSSPEELSSVPSSTIAVVRPQD 470  
DB 117 ---pqvatllidrfvkggrgqldkldltltafypgylcsspeelssvpsstiaavrpgd 172  
OY 471 IDTCRQDLVLYPKARLAFQNMNGSEYFVKIOSFLGAPREDIKALSOQNVSMDLATFM 530  
DB 173 idtcprqldvlypkarlaifqnmngseyfvkioqslfgapredikalsgqnvsmldatfm 232  
OY 531 KLRDVALPLTVAEVQKLLGPHVEGLKAEERHRRPVDMILRQRODDLTGLGIGGIPN 590  
DB 233 klrdvalplltvaevqkllgphveglkaeerhrpvrwmilrqrddldtllgiglggipn 292  
OY 591 GYLVIDLSVQ 600  
DB 293 gylvidlsvq 302  
  
RESULT 5  
ID AAB08543 standard; Protein: 399 AA.  
AC AAB08543;  
DT 20-DEC-2000 (first entry)  
XX  
DE Mesothelin related antigen (MRA)-1 polypeptide sequence.  
XX  
KW Mesothelin; mesothelin related antigen; MRA; malignant condition; cancer;  
KW differentiation antigen; adenocarcinoma; mesothelioma; ovarian carcinoma;  
KW pancreatic carcinoma; non-small cell lung carcinoma; MRA-1.  
XX  
OS Homo sapiens.  
XX  
PN WO200050900-A2.  
XX  
PD 31-AUG-2000.  
XX  
PF 25-FEB-2000; 2000WO-US04834.

XX  
PR 26-FEB-1999; 99US-0121767.  
PR 05-AUG-1999; 99US-0147404.  
XX  
PA (PACI-) PACIFIC NORTHWEST RES FOUND.  
XX  
PI Scholler NB, Hellstrom I, Hellstrom KE;  
XX  
DR WPI: 2000-572118/53.  
DR N-PSDB: AAA64362.  
XX  
PT Diagnosing carcinomas, e.g. adenocarcinoma or ovarian carcinomas, by  
PT detecting reactivity of a molecule (with an antigenic determinant and  
PT present in a sample) with an antibody specific for a mesothelin related  
PT antigen polypeptide  
XX  
PS Claim 48; Fig 5A-B; 92pp; English.  
XX  
CC The present sequence represents a human mesothelin related antigen  
CC (MRA)-1. Antibodies specific to MRA polypeptides are used for the  
CC detection of a malignant condition. Mesothelin is a differentiation  
CC antigen which is expressed in the surfaces of normal mesothelial cells  
CC and also on certain cancer cells, including epithelial ovarian tumours  
CC and mesotheliomas. The method is useful for detecting the presence of a  
CC malignant condition, specifically, adenocarcinoma, mesothelioma,  
CC ovarian carcinoma, pancreatic carcinoma or non-small cell lung carcinoma.  
XX  
SQ Sequence 399 AA;  
  
Query Match 46.9%; Score 1531; DB 21; Length 399;  
Best Local Similarity 97.1%; Pred. No. 2.2e-131;  
Matches 298; Conservative 1; Mismatches 0; Indels 8; Gaps 1;  
  
OY 294 EYEKTCACPSGKKAREIDSLFFKKWELEACVDALATQMDRVNAIPFTYEQDLYLKH 353  
DB 1 evektacpsgkkaareideslffkkweleacvdaallatqmdrvnaipfyeqldvllkh 60  
OY 354 LDELYPGYPSVSIQHLGYLFLKMSPEDIRKKNWTSLETLKALLEVDKGHEMSPOAPRR 413  
DB 61 ldelypgypsvsiqhlgyllfklmspedirknwtsletlkallevnkghems----- 113  
OY 414 LQVATLIDRFVKGKGGQLDKDLDLTLTAFYPGYLCSSPEELSSVPSSTIAVVRPQD 473  
DB 114 -pqvatllidrfvkggrgqldkldltltafypgylcsspeelssvpsstiaavrpgd 172  
OY 474 CDRQDLVLYPKARLAFQNMNGSEYFVKIOSFLGAPREDIKALSOQNVSMDLATFM 533  
DB 173 cdrqldvlypkarlaifqnmngseyfvkioqslfgapredikalsgqnvsmldatfm 232  
OY 534 TDAVPLTVAEVQKLLGPHVEGLKAEERHRRPVDMILRQRODDLTGLGIGGIPN 593  
DB 233 tdavplltvaevqkllgphveglkaeerhrpvrwmilrqrddldtllgiglggipngyl 292  
OY 594 VLDLSVQ 600  
DB 293 vldlsvq 299  
  
RESULT 6  
ID AAR53991 standard; peptide: 65 AA.  
AC AAR53991;  
DT 08-DEC-1994 (first entry)  
XX  
DE Meg-Pot fragment.  
XX  
KW Megakaryocyte potentiator; Meg-Pot; thrombocytopoiesis;  
KW platelet; amplification; primer; polymerase chain reaction; PCR.  
XX  
OS Homo sapiens.

```

XX MO9410312-A.
XX 11-MAY-1994.
XX 25-OCT-1993; 93WO-JP01540.
XX 23-OCT-1992; 92JP-0286153.
XX 11-NOV-1992; 92JP-0301387.
XX 09-DEC-1992; 92JP-0329546.
XX (CHUS ) CHUGAI SEIYAKU KK.
XX Hattori K, Kojima T, Oh-eda M, Yamaguchi N;
XX WPI: 1994-167467/20.
XX N-PSDB: AA063971.
XX New mega:karyocyte potentiator - for potential treatment of
XX thrombocytopenia
XX PS Disclosure: Page 51; 74pp; Japanese.
XX CC Use of the primers given in AA077809-10 in PCR resulted in the
XX fragments given in AA063971.
XX CC DNA encoding Meg-Pot has potential use in treatment of
XX thrombocytopenia and low platelet function.
XX SQ Sequence 65 AA:

```

```

Query Match 10.0%; Score 325; DB 15; Length 65;
Best Local Similarity 100.0%; Pred. No. 2.8e-22;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 64 LIGFPCAEEVSGISTEVRVLAVALAOKNVKLTSEQLRCIAHRLSEPPEDLDALPDLULLF 123
DB 1 llyfcaevsgistervrelavalaqknvklsteglrclahrlseppedaipdlillf 60
QY 124 LNPDA 128
DB 61 lnpda 65

```

```

RESULT 7
AAR53988
ID AAR53988 standard; peptide: 35 AA.
AC AAR53988;
XX 08-DEC-1994 (first entry)
DT Meg-Pot N-terminal peptide.
XX Meg-Pot N-terminal peptide.
DE Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;
XX Platelet.
KW Homo sapiens.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 19 /note="residue not identified"
XX MO9410312-A.
XX PD 11-MAY-1994.
XX PF 25-OCT-1993; 93WO-JP01540.
XX PR 23-OCT-1992; 92JP-0286153.
XX PR 11-NOV-1992; 92JP-0301387.
XX PR 09-DEC-1992; 92JP-0329546.
XX

```

```

PA (CHUS ) CHUGAI SEIYAKU KK.
XX Hattori K, Kojima T, Oh-eda M, Yamaguchi N;
XX WPI: 1994-167467/20.
XX New mega:karyocyte potentiator - for potential treatment of
XX thrombocytopenia
XX PS Disclosure: Page 49; 74pp; Japanese.
XX CC DNA encoding Meg-Pot has potential use in treatment of
XX thrombocytopenia and low platelet function.
XX SQ Sequence 35 AA:

```

```

Query Match 5.0%; Score 163; DB 15; Length 35;
Best Local Similarity 97.1%; Pred. No. 7.1e-08;
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 83 LAVAALOKNVKLTSEQLRCIAHRLSEPPEDLDALP 117
DB 1 lavalagknvklsteglrclahrlseppedaalp 35

```

```

RESULT 8
AAR53990
ID AAR53990 standard; peptide: 25 AA.
AC AAR53990;
XX 08-DEC-1994 (first entry)
DT Meg-Pot fragment.
XX Meg-Pot fragment.
DE Megakaryocyte potentiator; Meg-Pot; thrombocytopenia;
XX Platelet; amplification; primer: polymerase chain reaction; PCR.
XX KW Platelet; amplification; primer: polymerase chain reaction; PCR.
XX OS Homo sapiens.
XX MO9410312-A.
XX PD 11-MAY-1994.
XX PF 25-OCT-1993; 93WO-JP01540.
XX PR 23-OCT-1992; 92JP-0286153.
XX PR 11-NOV-1992; 92JP-0301387.
XX PR 09-DEC-1992; 92JP-0329546.
XX PA (CHUS ) CHUGAI SEIYAKU KK.
XX PI Hattori K, Kojima T, Oh-eda M, Yamaguchi N;
XX WPI: 1994-167467/20.
XX N-PSDB: AA063970.
XX PT New mega:karyocyte potentiator - for potential treatment of
XX thrombocytopenia
XX PS Disclosure: Page 50; 74pp; Japanese.
XX CC Use of the primers given in AA077805-08 in PCR resulted in the
XX fragments given in AA063969-70.
XX CC DNA encoding Meg-Pot has potential use in treatment of
XX thrombocytopenia and low platelet function.
XX SQ Sequence 25 AA:

```

```

Query Match 3.9%; Score 128; DB 15; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;

```

Matches	25; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
OY	106	LSPPEDALPLDLLFNPDAFS 130			
Db	1	lseppeidaipdlilfinpdafs 25			
RESULT	9				
AAE03600					
ID	AAE03600	standard; Protein; 794 AA.			
XX	AAE03600;				
AC					
XX					
DT	07-AUG-2001	(first entry)			
XX					
DE	Human	leucine-rich repeat-containing protein, AZAD.			
XX					
KW	Human;	leucine-rich repeat; AZAD; neurodegenerative disorder;			
KW	CNS disorder;	central nervous system disorder; prostate disorder;			
KW	prostatitis;	benign prostatic hyperplasia; adenocarcinoma;			
KW	prostate cancer;	genitourinary system carcinoma; testicular tumour;			
KW	Alzheimer's disease;	dementia; Parkinson's disease; multiple sclerosis;			
KW	amyotrophic lateral sclerosis;	epilepsy; psychiatric disorder; mania;			
KW	depression;	schizophrenia; anxiety; phobic disorder; learning disorder;			
KW	memory disorder;	amnesia; migraine; protein-protein interaction;			
KW	cellular activity;	neoplastic transformation; drug screening;			
KW	forensic identification;	gene therapy.			
XX					
OS	Homo sapiens.				
XX					
PH	Key	Location/Qualifiers			
FT	Peptide	1..58			
FT		/label= Signal_peptide			
FT	Protein	59..794			
FT		/label= Mature_AZAD_protein			
FT	Region	65..94			
FT		/note= "N-terminal leucine-rich repeat (LRR)"			
FT	Region	427..456			
FT		/note= "N-terminal leucine-rich repeat (LRR)"			
FT	Region	96..119			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	120..143			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	144..167			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	168..191			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	192..215			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	216..239			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	240..263			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	264..287			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	288..311			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	312..333			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	458..481			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	482..505			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	506..529			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	530..553			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	554..577			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	578..601			
FT		/note= "Leucine-rich repeat (LRR)"			
FT	Region	602..625			
FT		/note= "Leucine-rich repeat (LRR)"			

FT	Region	626..649	
FT		/note= "Leucine-rich repeat (LRR)"	
FT	Region	651..674	
FT		/note= "Leucine-rich repeat (LRR)"	
FT	Region	676..697	
FT		/note= "Leucine-rich repeat (LRR)"	
FT	Region	698..719	
FT		/note= "Leucine-rich repeat (LRR)"	
FT	Region	707..755	
FT		/note= "C-terminal leucine-rich repeat (LRR)"	
XX			
XX	WO200142286-A2.		
PN			
PD	14-JUN-2001.		
XX			
PF	07-DEC-2000; 2000WO-US33140.		
XX			
PR	08-DEC-1999; 99US-0456592.		
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
PI	Khodadoust MM;		
XX			
DR	WPI: 2001-381633/40.		
XX			
DR	N-PSDB: AAD08044.		
XX			
PT	New leucine-rich repeat-containing protein and nucleic acid molecules		
PT	for diagnosing, treating neural disorders, such as neurodegenerative		
PT	disorders, such as Alzheimer's disease, dementia, epilepsy and prostate		
PT	cancer		
XX			
PS	Claim 12; Fig 2; 133p; English.		
XX			
CC	The present sequence is human leucine-rich repeat-containing		
CC	secreted protein, AZAD. AZAD polypeptide is useful for		
CC	identifying a compound which modulates its activity and binds to it.		
CC	AZAD proteins and nucleic acid molecules are useful for treating and		
CC	diagnosing AZAD-mediated or related disorders, which includes a neural		
CC	disorder (e.g. neurodegenerative disorders, including CNS disorders)		
CC	and a prostate disorder, e.g. prostatitis, benign prostatic hyperplasia		
CC	and cancer (e.g. adenocarcinoma, prostate cancer, genitourinary system		
CC	carcinomas and testicular tumour). The neurodegenerative disorders		
CC	include Alzheimer's disease, dementias related to Alzheimer's disease,		
CC	Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis,		
CC	epilepsy, psychiatric disorders e.g. depression, schizophrenia		
CC	disorders, mania, anxiety or phobic disorders, learning or memory		
CC	disorders e.g. amnesia or age-related memory loss, and neurological		
CC	disorders e.g. migraine. AZAD molecules are useful as markers of		
CC	for predisposition of disease state or as markers of drug activity or		
CC	pharmacogenomic profile of a subject. The AZAD polypeptide and		
CC	polynucleotide are capable of modulating protein-protein interaction,		
CC	e.g. by interacting with an extracellular component, thereby modulating		
CC	cellular activities, including attachment, adhesion, migration,		
CC	patterning, growth and/or differentiation of a cell. AZAD proteins		
CC	regulate embryonic development and differentiation, tissue maintenance		
CC	and function, pathological conditions, e.g. neuronal degeneration,		
CC	neoplastic transformation and tumour progression. AZAD proteins are		
CC	useful as immunogens to raise anti-AZAD antibodies which are useful to		
CC	detect and isolate AZAD proteins and modulate its activity. AZAD		
CC	proteins are useful to screen for naturally occurring AZAD substrates		
CC	and to screen for drugs or compounds which modulate AZAD activity.		
CC	AZAD nucleic acid fragments are useful as primers or hybridisation		
CC	probes for the detection of AZAD-encoding nucleic acids. cDNA encoding		
CC	the AZAD protein are useful in gene therapy. AZAD sequences are also		
CC	useful to map their respective genes on a chromosome, for tissue typing		
CC	and in forensic identification of a biological sample.		
XX			
XX	Sequence	794 AA;	

Query Match 3.8%; Score 123.5; DB 22; Length 794;  
Best Local Similarity 29.0%; Pred. No. 0.037;

Matches 72; Conservative 28; Mismatches 107; Indels 41; Gaps 15;

QY 15 RPSGL--LFLFSLGWHPARTLAGETESAPLGCVITTPHN--ISSLSPPQLGFPQAE 71  
DB 183 rpgtfgalgaatclnlnalnalylpamafggllrvrwlrlshnalsvlapaalaglpalr 242  
QY 72 VSGSTEREHLAVALAOKNKVLTSEQLRCLAH--RLSEPP-----EDLDLP--LDLTL 122  
DB 243 rlslhnelgqlppv-----lsqarqarlelghnpltyageedgialpglirell 294  
QY 123 -----FLNPAESGPOACTFEFSRITKAN--VDLIP--RGAPERQLLPALACGW--R 171  
DB 295 dggalqalgaipraf--hcprlhtldlrngldtclpdpqpgqlrrlrqgircgaaar 351  
QY 172 G-SLSEADVRALGGLACDLP--RFVAESAELLPRVSCGRLDQDQEAARAALQG- 227  
DB 352 gplrvagagaaalgr---rvpgprtlrgealdalrpdwlrpcgdaageeileerava 408  
QY 228 GGPYPGP 235  
DB 409 rapprgpp 416

## RESULT 10

AAB82352 standard; Protein: 794 AA.

AAB82352;

23-JUL-2001 (first entry)

Protein sequence SEQ ID NO.2.

JAFFA; human; fibroblast growth factor.

Homo sapiens.

WO200138357-A2.

31-MAY-2001.

22-NOV-2000; 2000WO-US32181.

22-NOV-1999; 99US-0444165.

(MILL-) MILLENNIUM PHARM INC.

Khodadoust MM;

WPI; 2001-355881/37.

N-PSDB; AAF90331, AAF90332.

Isolated JAFFA nucleic acid molecules which encode novel fibroblast growth factor family members (JAFFA) are useful for developing novel diagnostic and therapeutic agents for JAFFA-associated disorders such as cancers

Disclosure; Page 129-131; 137pp; English.

The present invention relates to human JAFFA polynucleotides (see AAF90326) and polypeptides (see AAB82351). JAFFA is a novel member of the fibroblast growth factor family. JAFFA proteins may be useful for developing novel diagnostic and therapeutic agents for JAFFA-associated disorders such as cancers and ABO(H) blood group disorders, and for controlling cellular proliferative and/or differentiative disorders. JAFFA nucleic acids and proteins may be used to treat and/or diagnose a variety of immune disorders such as autoimmune disease and multiple sclerosis. JAFFA nucleic acids may also be used to express JAFFA protein, to detect JAFFA mRNA or a genetic alteration in a JAFFA gene, and to modulate JAFFA activity. Note: The present sequence is given in the Sequence Listing from the present invention, but does not correspond with the SEQ ID NO. described within the specification.

XX Sequence 794 AA;

Query Match 3.88; Score 123.5; DB 22; Length 794;  
Best Local Similarity 29.08; Pred. No. 0.037;  
Matches 72; Conservative 28; Mismatches 107; Indels 41; Gaps 15;

QY 15 RPSGL--LFLFSLGWHPARTLAGETESAPLGCVITTPHN--ISSLSPPQLGFPQAE 71  
DB 183 rpgtfgalgaatclnlnalnalylpamafggllrvrwlrlshnalsvlapaalaglpalr 242  
QY 72 VSGSTEREHLAVALAOKNKVLTSEQLRCLAH--RLSEPP-----EDLDLP--LDLTL 122  
DB 243 rlslhnelgqlppv-----lsqarqarlelghnpltyageedgialpglirell 294  
QY 123 -----FLNPAESGPOACTFEFSRITKAN--VDLIP--RGAPERQLLPALACGW--R 171  
DB 295 dggalqalgaipraf--hcprlhtldlrngldtclpdpqpgqlrrlrqgircgaaar 351  
QY 172 G-SLSEADVRALGGLACDLP--RFVAESAELLPRVSCGRLDQDQEAARAALQG- 227  
DB 352 gplrvagagaaalgr---rvpgprtlrgealdalrpdwlrpcgdaageeileerava 408  
QY 228 GGPYPGP 235  
DB 409 rapprgpp 416

## RESULT 11

AAG91445 standard; Protein: 943 AA.

AAG91445;

26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 5199.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

organic acid synthesis.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

N-PSDB; AAH66664.

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

Claim 17; SEQ ID NO: 5199; 246pp + sequence listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and





```
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155179.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159564.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.58; Score 114.5; DB 21; Length 349;
Best Local Similarity 20.78; Pred. No. 0.069;
Matches 92; Conservative 63; Mismatches 128; Indels 161; Gaps 23;

QY 79 RVREIYALAKQNVKLTSTQRLCIA--HRISEPPEDLDALPLDLLFLNPDAFSGPOACT 136
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 rvrvm--klgsqglevaeagqcmalserygapkpetadial-----lhainsg-----v 54

QY 137 RFESEIRIKANVDLPKCAPERORLLPAAACGVR-----GSLISEADVRAAGCLA 187
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 tff-----dstadmygpecneillgkalkdykvelatkigffivegeisevirg-- 104

QY 188 CDLPGRFVAESAEEVLLPPL-VSCPG-----PLDDQDQGEAARALOGG-----GPP 231
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 ---dpeyvraaceaslkrldiacidlyyqhndrcvpleimrelkklveegkiklygls 161

QY 232 YGPST-----WSVSTWDALRGLLPV---LGOPILRSIP--QGIVAAWR 270
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 easastlrhavnpitavqlewsrdaeedlplrcreljglvayspjrgflaagp 221

QY 271 Q--RSSRDPSSWRQPERTILRRFRREVEKTIACPSGKKAREIDESLIRYKK---WELBAC 324
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 klaenlendfirkattll--prfge-----nvdhnhkklfeksamaekgyc 266
```

QY 325 VDALLAT----QMDRNAIPFTYEQOLDVLRKHKLDELPGYPSVIOHGLYFLKMSPE 380  
 Db 267 tpaqalawvhqgdgdepigtk-----ienlgnitalavsklpe 309  
 QY 381 DIRKNVTSLETLKALEVDKGHEKSPQAPRRPRLPOVATFLDRFYKGRGQLDKDTLTTLT 440  
 Db 310 ei-----seldsl-----akpes-----vkge----- 326  
 QY 441 AFYEGYLCSLPELSSVPSISW 464  
 Db 327 ----rymasmtfknsucplissw 346  
 RESULT 13  
 AAG38204  
 ID AAG38204 standard; Protein: 351 AA.  
 AC AAG38204;  
 DT 18-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 47098.  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 OS Arabidopsis thaliana.  
 FN EPI033405-A2.  
 PD 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139839.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 21-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145099.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.



CC fragments given in AA063969-70.  
CC DNA encoding Meg-Pot has potential use in treatment of  
CC thrombocytopenia and low platelet function.  
XX  
SQ Sequence 23 AA:

Query Match 3.5%; Score 114; DB 15; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 ROLGFPCEVSGSLSTERYELA 84  
Db 1 RGLGFPCEVSGSLSTERYELA 23

## RESULT 15

AA051557  
ID AA051557 standard; Protein; 913 AA.

XX AA051557;

XX 18-MAY-2000 (first entry)

DE Human PLA2 protein.

XX PLA2; phospholipase A2; phosphatide 2-acyl hydrolase; human; therapy;  
KW arachidonic acid; lysophospholipid; Alzheimer's disease.

XX Homo sapiens.

XX US6025178-A.

PN 15-FEB-2000.

XX 28-MAR-1997; 97US-0827208.

XX 29-MAR-1996; 96US-0014608.

XX (ELIL ) LILLY & CO ELI.

PI Sharp JD, Scriffler BA, Chou XG, Kramer RM, Pickard RT;

DR WPI: 2000-181816/16.

DR N-PSDB; AA288756; AA288757.

PT An isolated amino acid having phospholipase (PL)A2 activity is useful  
PT in assays to identify inhibitors having a therapeutic benefit, such as  
PT inhibiting the central role of PLA2 in the inflammatory component of  
PT Alzheimer's disease -

PS Claim 1; Column 53-58; 32pp; English.

XX This invention describes a novel human phospholipase A2 (PLA2) protein  
CC (1) and its encoding nucleic acid. The amino acid (1) releases  
CC arachidonic acid in specific tissues characterized by unique membrane  
CC phospholipids, by generating lysophospholipid species which are  
CC deleterious to membrane integrity or by remodeling of unsaturated species  
CC of membrane phospholipids through deacylation/reacylation mechanisms. The  
CC amino acid is useful in assays to identify inhibitors having a  
CC therapeutic benefit, such as inhibiting the central role of PLA2 in the  
CC inflammatory component of Alzheimer's disease. The amino acid (1) allows  
CC sensitive and rapid screening and identification of inhibitors of  
CC phospholipase A2. This sequence represents the human PLA2 protein (also  
CC known as phosphatide 2-acyl hydrolase).

XX Sequence 913 AA;

Query Match 3.5%; Score 113.5; DB 21; Length 913;  
Best Local Similarity 20.8%; Pred. No. 0.38;  
Matches 143; Conservative 72; Mismatches 257; Indels 215; Gaps 33;

OY 1 MALORLDPCCWSCGRPSGLFLFLFSLAGWVHPARTLAGETGES---APLG-GVLTTPHNI 56  
Db melkvfdqdlvvgddp---vlsvlfaqlr-----agefirresfslpsgggrlevefrl 260  
OY 57 SLSRPLLGPCEVSGSLSTERYELAVALAQNKKVSLSTEDLRCLARLSEPPEDDAL 116  
Db gsladr-----gewlvs---ngvlvarelscshvqleatgqkxse 298  
OY 117 PLDLFLFNPDAFSGPQACT-----RFE-SRTTANVDLLPRGAPREORLLPALAOWG 169  
Db hrvgllv---psgepqaasvgtgfrfhpcacwqealstlqdaeeqlkaplsalpsg 356  
OY 170 -----VRGSLSEADVR---AUGSLACDLPGRFVESAVAL---LPRLV 208  
Db gvvrlvftsgqplmrvtelkkaegrlrelavrlgfgpcaeeqafslrtkvvaalrlqql 416  
OY 209 CPGLPDODQEARALDGGPRPGPSTWSYSTDALRGLLPVLGQ---PIRISIQG 264  
Db 1dgdldedeiprvaimatggy-----lramtslygqlaglkqlldcvsyltga 466  
OY 265 IVAAMRORS-SHDPSPMRPERTILRRFRREVEKTCPSGKKARIDSLIFYKKWLEA 323  
Db sgstwalanlyedpewsqkd-----laqpt----- 491  
OY 324 CVDAALLATQMDRVNAPFTYEQDLVKKLDELVPQGPESVIOHLYFLKMSPEDIR 383  
Db 492 ---eliktgyknklyglapsqldgyrgelaerarlgypsclt----- 531  
OY 384 KWNVTSLETIKALLEVDGHEKSPQAF-----RPLPQVATLIDRFVKGROLDKDTLD 438  
Db 532 -nlwallnealldhepbhklisdgrealsbgqnlplycal-----ntkqg-----s 577  
OY 439 LTAEPGYLCSLSPBELSSVPPSSIMAVRPQDLDTCD--PROLDVLYPKARLAF-ONNG 495  
Db 578 lttfegwcefspeyvgf---pkysgfipseifsgelfmgqlmkrlpsesticflegws 634  
OY 496 SEYFVKIOSFL--GGAPTE--DLKALSOQVNSMDIATMKLR-----TDAVL 538  
Db 635 nlyaanldqsllywaspepfdwrvrnqanldkeqpilkieepstagraiaefftdilt 694  
OY 539 --PLTVAEVOKLPGHVEGLKAEERHRRPVBDWILRQDDLDLTGSLGGGIPN----- 590  
Db 695 wrplagathnflrghfn--kdyfqhpnfstw---katlid-----gipnqltpse 740  
OY 591 -----GYLVLDLSVOETLSGPCL 609  
Db 741 phclldvgyll-----ntscsl 757

Search completed: December 7, 2001, 10:01:37  
Job time: 720 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 7, 2001, 09:48:52 ; Search time 22.61 Seconds  
(Without alignments)  
625.036 Million cell updates/sec

Title: US-09-215-035-2

Perfect score: 3261  
Sequence: 1 MALGRDPCWCGCDRPGSL.....LLGPGVLTVALLASTIA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3261	100.0	628	3 US-08-776-271-2	Sequence 2, Appl1
2	3261	100.0	628	4 US-09-215-035-2	Sequence 2, Appl1
3	3055	93.7	622	1 US-08-426-819A-35	Sequence 35, Appl1
4	2956	90.6	584	1 US-08-426-819A-37	Sequence 36, Appl1
5	1247	38.2	248	1 US-08-426-819A-36	Sequence 37, Appl1
6	325	10.0	65	1 US-08-426-819A-33	Sequence 33, Appl1
7	210	5.0	40	1 US-08-426-819A-19	Sequence 19, Appl1
8	163	5.0	35	1 US-08-426-819A-5	Sequence 5, Appl1
9	134	4.1	6095	4 US-09-144-085-2	Sequence 2, Appl1
10	113.5	3.5	778	4 US-09-460-145-2	Sequence 2, Appl1
11	113.5	3.5	797	4 US-08-460-145-4	Sequence 2, Appl1
12	113.5	3.5	913	3 US-08-827-208-3	Sequence 3, Appl1
13	113.5	3.5	913	4 US-09-500-358-3	Sequence 3, Appl1
14	113.5	3.5	913	3 US-09-498-809-3	Sequence 3, Appl1
15	111	3.4	33	1 US-08-426-819A-18	Sequence 18, Appl1
16	107.5	3.3	1288	3 US-08-762-428A-6	Sequence 6, Appl1
17	107.5	3.3	3567	2 US-07-642-734C-4	Sequence 4, Appl1
18	107.5	3.3	3567	3 US-08-439-009A-4	Sequence 4, Appl1
19	106.5	3.3	634	1 US-08-338-152A-17	Sequence 17, Appl1
20	106.5	3.3	653	1 US-08-339-152A-16	Sequence 16, Appl1
21	106.5	3.3	653	2 US-08-007-999B-3	Sequence 3, Appl1
22	106.5	3.3	653	2 US-08-689-276A-3	Sequence 3, Appl1
23	105	3.2	1239	1 US-08-026-138E-3	Sequence 3, Appl1
24	103.5	3.2	834	5 PCT-US95-14418-2	Sequence 2, Appl1
25	103.5	3.2	834	5 PCT-US95-15327-2	Sequence 2, Appl1
26	101.5	3.1	834	1 US-07-977-434-10	Sequence 10, Appl1
27	101.5	3.1	834	1 US-08-073-384C-6	Sequence 6, Appl1

28	101.5	3.1	834	1 US-08-254-359A-6	Sequence 6, Appl1
29	101.5	3.1	834	1 US-08-384-490-31	Sequence 31, Appl1
30	101.5	3.1	834	1 US-08-483-043-6	Sequence 6, Appl1
31	101.5	3.1	834	1 US-08-459-383-31	Sequence 31, Appl1
32	101.5	3.1	834	1 US-08-458-819-10	Sequence 10, Appl1
33	101.5	3.1	834	1 US-08-481-238-6	Sequence 6, Appl1
34	101.5	3.1	834	2 US-08-471-066B-6	Sequence 6, Appl1
35	101.5	3.1	834	2 US-08-484-956-6	Sequence 6, Appl1
36	101.5	3.1	834	2 US-08-757-653-6	Sequence 6, Appl1
37	101.5	3.1	834	2 US-08-589-491-6	Sequence 6, Appl1
38	101.5	3.1	834	2 US-08-756-386-6	Sequence 6, Appl1
39	101.5	3.1	834	2 US-08-823-516-6	Sequence 6, Appl1
40	101.5	3.1	834	3 US-08-682-853A-6	Sequence 6, Appl1
41	101.5	3.1	834	3 US-08-759-038-6	Sequence 6, Appl1
42	101.5	3.1	834	3 US-08-758-314-6	Sequence 6, Appl1
43	101.5	3.1	834	5 PCT-US91-07035-10	Sequence 10, Appl1
44	100.5	3.1	609	4 US-08-980-115-11	Sequence 11, Appl1
45	99.5	3.1	514	4 US-09-385-028-5	Sequence 5, Appl1

## ALIGNMENTS

RESULT 1  
US-08-776-271-2  
Sequence 2, Application US/08776271  
Patent No. 6083502  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
TITLE OF INVENTION: Mesothelin, a Differentiation Antigen  
TITLE OF INVENTION: Present on Mesothelioma, Mesotheliomas and Ovarian Cancers  
TITLE OF INVENTION: and Methods and Kits for Targeting the Antigen  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,271  
FILING DATE: 01-DEC-1998  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/00224  
FILING DATE: 03-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/010,166  
FILING DATE: 05-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Paris, Susan K. 41,739  
REGISTRATION NUMBER: 015280-25910005  
REFERENCE/DOCKET NUMBER: 015280-25910005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 628 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-776-271-2  
Query Match 100.0%; Score 3261; DB 3; Length 628;

Best Local Similarity 100.0%; Pred. No. 1.5e-303;  
Matches 628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MALQRLDPCWSCGDRGSLFLFLSLGWHPARTLAGETGESAPLGGVLTTPHNISLS 60
DB 1 MALQRLDPCWSCGDRGSLFLFLSLGWHPARTLAGETGESAPLGGVLTTPHNISLS 60
OY 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNVKSTEDQLCLAHRLSEPPEDLDALPDL 120
DB 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNVKSTEDQLCLAHRLSEPPEDLDALPDL 120
OY 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNVKSTEDQLCLAHRLSEPPEDLDALPDL 120
DB 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNVKSTEDQLCLAHRLSEPPEDLDALPDL 120
OY 121 LFLNPDASFSGQACTRFSRTTKANVLLPRGAPERQRLPALACWVSGSLSEADY 180
DB 121 LFLNPDASFSGQACTRFSRTTKANVLLPRGAPERQRLPALACWVSGSLSEADY 180
OY 181 RALGGLACDLPRFVAESAENVLLPRLVSCPGPLDDQDEAARAALOGGPPYGPSTWSV 240
DB 181 RALGGLACDLPRFVAESAENVLLPRLVSCPGPLDDQDEAARAALOGGPPYGPSTWSV 240
OY 241 STMALRGLLPVLPQPIIRSIPOGIVAAMRORSSRDPSPROBERTILRRFRREVEKTAC 300
DB 241 STMALRGLLPVLPQPIIRSIPOGIVAAMRORSSRDPSPROBERTILRRFRREVEKTAC 300
OY 301 PSGKAREIDESLIFFKKWELEACVDALLATQMDRVNAIPFTYQDLVYLKHKDELTPQ 360
DB 301 PSGKAREIDESLIFFKKWELEACVDALLATQMDRVNAIPFTYQDLVYLKHKDELTPQ 360
OY 361 GYPSVIOHLGLFLKMSPEDIRKNNVTSLETALKALEVDKCHEMSPOAPRRRLPOQVATL 420
DB 361 GYPSVIOHLGLFLKMSPEDIRKNNVTSLETALKALEVDKCHEMSPOAPRRRLPOQVATL 420
OY 421 IDRFYKRGQDLKDTLDLTAFAPGYLCSLSPEELSSVPSISMAVRQDDLTCDPRQD 480
DB 421 IDRFYKRGQDLKDTLDLTAFAPGYLCSLSPEELSSVPSISMAVRQDDLTCDPRQD 480
OY 481 VLYPKARLAFQNMNGSEFFVATIOSFLGAPTEDLKALSOQVNSMDLAFMKLRTDAVPL 540
DB 481 VLYPKARLAFQNMNGSEFFVATIOSFLGAPTEDLKALSOQVNSMDLAFMKLRTDAVPL 540
OY 541 TVAEYQKLLGPHVEGLKAEERHPRVDMILRODODDLTGLGOGGIPNGYLVLDLSVQ 600
DB 541 TVAEYQKLLGPHVEGLKAEERHPRVDMILRODODDLTGLGOGGIPNGYLVLDLSVQ 600
OY 601 ETLGTPCLLGGPVLYVLAALLASTLA 628
DB 601 ETLGTPCLLGGPVLYVLAALLASTLA 628

RESULT 2
US-09-215-035-2
; Sequence 2, Application US/09215035
; Patent No. 6153430
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; TITLE OF INVENTION: Mesothelin, a Differentiation Antigen
; TITLE OF INVENTION: Present on Mesothelium, Mesotheliomas and Ovarian Cancers
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215,035

```

```

; FILING DATE: No. 6153430 yet assigned
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/776,271
; FILING DATE: 01-DEC-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/00224
; FILING DATE: 03-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/010,166
; FILING DATE: 05-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fails, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-259110US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-215-035-2

Query Match 100.0%; Score 3261; DB 4; Length 628;
Best Local Similarity 100.0%; Pred. No. 1.5e-303;
Matches 628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MALQRLDPCWSCGDRGSLFLFLSLGWHPARTLAGETGESAPLGGVLTTPHNISLS 60
DB 1 MALQRLDPCWSCGDRGSLFLFLSLGWHPARTLAGETGESAPLGGVLTTPHNISLS 60
OY 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNVKSTEDQLCLAHRLSEPPEDLDALPDL 120
DB 61 PROLLGFCAEVSGSLSTERVRELAVALAOKNVKSTEDQLCLAHRLSEPPEDLDALPDL 120
OY 121 LFLNPDASFSGQACTRFSRTTKANVLLPRGAPERQRLPALACWVSGSLSEADY 180
DB 121 LFLNPDASFSGQACTRFSRTTKANVLLPRGAPERQRLPALACWVSGSLSEADY 180
OY 181 RALGGLACDLPRFVAESAENVLLPRLVSCPGPLDDQDEAARAALOGGPPYGPSTWSV 240
DB 181 RALGGLACDLPRFVAESAENVLLPRLVSCPGPLDDQDEAARAALOGGPPYGPSTWSV 240
OY 241 STMALRGLLPVLPQPIIRSIPOGIVAAMRORSSRDPSPROBERTILRRFRREVEKTAC 300
DB 241 STMALRGLLPVLPQPIIRSIPOGIVAAMRORSSRDPSPROBERTILRRFRREVEKTAC 300
OY 301 PSGKAREIDESLIFFKKWELEACVDALLATQMDRVNAIPFTYQDLVYLKHKDELTPQ 360
DB 301 PSGKAREIDESLIFFKKWELEACVDALLATQMDRVNAIPFTYQDLVYLKHKDELTPQ 360
OY 361 GYPSVIOHLGLFLKMSPEDIRKNNVTSLETALKALEVDKCHEMSPOAPRRRLPOQVATL 420
DB 361 GYPSVIOHLGLFLKMSPEDIRKNNVTSLETALKALEVDKCHEMSPOAPRRRLPOQVATL 420
OY 421 IDRFYKRGQDLKDTLDLTAFAPGYLCSLSPEELSSVPSISMAVRQDDLTCDPRQD 480
DB 421 IDRFYKRGQDLKDTLDLTAFAPGYLCSLSPEELSSVPSISMAVRQDDLTCDPRQD 480
OY 481 VLYPKARLAFQNMNGSEFFVATIOSFLGAPTEDLKALSOQVNSMDLAFMKLRTDAVPL 540
DB 481 VLYPKARLAFQNMNGSEFFVATIOSFLGAPTEDLKALSOQVNSMDLAFMKLRTDAVPL 540
OY 541 TVAEYQKLLGPHVEGLKAEERHPRVDMILRODODDLTGLGOGGIPNGYLVLDLSVQ 600
DB 541 TVAEYQKLLGPHVEGLKAEERHPRVDMILRODODDLTGLGOGGIPNGYLVLDLSVQ 600
OY 601 ETLGTPCLLGGPVLYVLAALLASTLA 628
DB 601 ETLGTPCLLGGPVLYVLAALLASTLA 628

```

Db 601 ETLSTGTPCLLGPVLTVALALLASTLA 628

RESULT 3

US-08-426-819A-35

Sequence 35, Application US/08426819A

Patent No. 5723318

GENERAL INFORMATION:

APPLICANT: Yamaguchi, No. 5723318oml

APPLICANT: Kojima, Tetsuo

APPLICANT: Oh-Eda, Masayoshi

APPLICANT: Hatcori, Kunihiro

TITLE OF INVENTION: Genes Coding for Megakaryocyte Potentiator

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/426,819A

FILING DATE: 21-Apr-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 230-107P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 622 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-426-819A-35

Query Match 93.78; Score 3055; DB 1; Length 622;

Best Local Similarity 93.48; Pred. No. 7.8e-284;

Matches 601; Conservative 3; Mismatches 16; Indels 10; Gaps 3;

QY 1 MALRLDP-CNSGDRP-GSLFLFSLGAVHPARTLAGETGESAPLGVLTTPPHNISS 58

Db 1 MALTPARILGSCGTPALGSLFLFSLGAVHPARTLAGETGESAPLGVLTTPPHNISS 60

QY 59 LSPROLLGFPCEAVESGLSTERYRELAVALAOKNKLSTEQRLCLAHRLSEPPEDDALPL 118

Db 61 LSPROLLGFPCEAVESGLSTERYRELAVALAOKNKLSTEQRLCLAHRLSEPPEDDALPL 120

QY 119 DLLFLFNPDASGQACRFRTSKANVDLLPRGAPRQRLTPALACNVRGSLSEA 178

Db 121 DLLFLFNPDASGQACRFRTSKANVDLLPRGAPRQRLTPALACNVRGSLSEA 180

QY 179 DYRALGGLACDLPGRFVASEVLLPRLVSCPGPLDDQOQANARALGGGPPGPPSTW 238

Db 181 DYRALGGLACDLPGRFVASEVLLPRLVSCPGPLDDQOQANARALGGGPPGPPSTW 240

QY 239 SVSTMDALRGLLPLVGOPIIRSIPOGIVAANRORSSRPMSROBERTILRPRFREVEKT 298

Db 241 SVSTMDALRGLLPLVGOPIIRSIPOGIVAANRORSSRPMSROBERTILRPRFREVEKT 300

QY 299 ACPGSKAREIDESLIFPKKWEACVDALLATOMDRVNAIPTTYEQLDVKKHLDLY 358

Db 301 ACPGSKAREIDESLIFPKKWEACVDALLATOMDRVNAIPTTYEQLDVKKHLDLY 360

QY 359 POGYPESVIOHLYFLFKMSPEDIRKMNVTSLFTKALLEVDKHEMSPOARRPLPOVA 418

Db 361 POGYPESVIOHLYFLFKMSPEDIRKMNVTSLFTKALLEVDKHEMSPOARRPLPOVA 412

QY 419 TLIDRFVGRQDLKDDTLDTLTAFFPGYLCSLSPEDLSSVPSSIMAVRPDDLTCDPRQ 478

Db 413 TLIDRFVGRQDLKDDTLDTLTAFFPGYLCSLSPEDLSSVPSSIMAVRPDDLTCDPRQ 472

QY 479 LDVLYPKARLAFQNNNGSEYVKIOSFTGAPTEDLKALSOONVSMIDLATPMKLRTPDAVL 538

Db 473 LDVLYPKARLAFQNNNGSEYVKIOSFTGAPTEDLKALSOONVSMIDLATPMKLRTPDAVL 532

QY 539 PLTYAEVOKLLGPHVEGKAEERRRPVMDWTLRQDDDLTGIGLOGINGVLYLDLS 598

Db 533 PLTYAEVOKLLGPHVEGKAEERRRPVMDWTLRQDDDLTGIGLOGINGVLYLDLS 592

QY 599 VOETLSTGTPCLLGPVLTVALALLASTLA 628

Db 593 VOETLSTGTPCLLGPVLTVALALLASTLA 622

RESULT 4

US-08-426-819A-36

Sequence 36, Application US/08426819A

Patent No. 5723318

GENERAL INFORMATION:

APPLICANT: Yamaguchi, No. 5723318oml

APPLICANT: Kojima, Tetsuo

APPLICANT: Oh-Eda, Masayoshi

APPLICANT: Hatcori, Kunihiro

TITLE OF INVENTION: Genes Coding for Megakaryocyte Potentiator

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/426,819A

FILING DATE: 21-Apr-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 230-107P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 584 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: internal

US-08-426-819A-36

Query Match 90.68; Score 2956; DB 1; Length 584;

Best Local Similarity 97.38; Pred. No. 2.1e-274;

Matches 576; Conservative 2; Mismatches 6; Indels 8; Gaps 1;





TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 65 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-426-819A-33

Query Match 10.0%; Score 325; DB 1; Length 65;  
Best Local Similarity 100.0%; Pred. No. 5.4e-24;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 LIGPCAEVSGLSITRVELVALAOKNVKISTEQLRLARLSPPDDALPLDLLF 123  
|||||  
DB 1 LIGPCAEVSGLSITRVELVALAOKNVKISTEQLRLARLSPPDDALPLDLLF 60  
|||||  
OY 124 LNPPA 128  
|||||  
DB 61 LNPPA 65

RESULT 7  
US-08-426-819A-19  
Sequence 19, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318om1  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiko  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
TITLE OF INVENTION: Potentilator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
IMMEDIATE SOURCE:  
CLONE: HPC15  
FEATURE:  
NAME/KEY: Peptide

LOCATION: 1..40 /label= fragment  
OTHER INFORMATION: /note= "sequence of cyanogen bromide fragment of MegPOT fr  
OTHER INFORMATION: CDNA in Table 3"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 32  
OTHER INFORMATION: /product= "is Xaa in COS"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 35  
OTHER INFORMATION: /product= "is Xaa in COS"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 37  
OTHER INFORMATION: /product= "is Xaa in COS"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..16 /label= fragment  
OTHER INFORMATION: /note= "sequence of cyanogen bromide fragment representing  
OTHER INFORMATION: C-terminus of MegPOT determined in Example 28"  
US-08-426-819A-19

Query Match 6.4%; Score 210; DB 1; Length 40;  
Best Local Similarity 100.0%; Pred. No. 2.5e-13;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 244 DALRGILPVGPIIRSIPOGIVAAWRSSRDPSPWROPE 283  
|||||  
DB 1 DALRGILPVGPIIRSIPOGIVAAWRSSRDPSPWROPE 40  
|||||

RESULT 8  
US-08-426-819A-5  
Sequence 5, Application US/08426819A  
Patent No. 5723318  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, No. 5723318om1  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiko  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
TITLE OF INVENTION: Potentilator  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819A  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid

STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
FRAGMENT TYPE: Internal  
IMMEDIATE SOURCE:  
CLONE: HPCYS  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..35  
OTHER INFORMATION: /label- peptide  
OTHER INFORMATION: /note- "sequence of a Glu-C peptide of Megpor"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 6..33  
OTHER INFORMATION: /label- peptide  
OTHER INFORMATION: /note- "region of Glu-C peptide having low frequency of usage"  
US-08-426-819A-5

Query Match 5.0%; Score 163; DB 1; Length 35;  
Best Local Similarity 97.1%; Pred. No. 6.4e-09;  
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 83 LAVAIAOKNVKLTSEOLCLARLSEPPEDDLP 117  
DB 1 LAVAIAOKNVKLTSEOLCLARLSEPPEDDLP 35

RESULT 9  
US-09-144-085-2  
Sequence 2, Application US/09144085  
Patent No. 6280999  
GENERAL INFORMATION:  
APPLICANT: Gustafsson, Claes  
APPLICANT: Bellach, Mary C.  
APPLICANT: Ashley, Gary  
APPLICANT: Julien, Bryan  
APPLICANT: Ziermann, Rainer  
TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
FILE REFERENCE: 30062-20020.20  
CURRENT APPLICATION NUMBER: US/09/144,085  
CURRENT FILING DATE: 1998-08-31  
EARLIER APPLICATION NUMBER: 09/010,809  
EARLIER FILING DATE: 1998-01-22  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 6095  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-144-085-2

Query Match 4.1%; Score 134; DB 4; Length 6095;  
Best Local Similarity 23.6%; Pred. No. 0.019;  
Matches 143; Conservative 64; Mismatches 217; Indels 182; Gaps 35;

OY 13 GDRPSLLFLFLSLGWN--HRAKTLAGETGTSAPLGVLTTPHHISSISPOL----- 64  
DB 2975 GMRPGVTALITGGVGLGHLARWLGR-GAEHL-----VLASRGCASAPGASLDELVA 3029  
OY 65 ---LGFPCAESGLSTERVRELAVALAOKNVKLTSEOLCLARL-----RLSE-DPE 111  
DB 3030 RGIRTLACDVS---ERA-QLAALLAE--LEODEAPLRAVAHLAGIGRRVPLELEBE 3082  
OY 112 DLD-----ALPLDLLFLNP-DAF-----SGPOACTRFFSRITKANVDLL 150  
DB 3083 QLEGLAKVKGAMHLHQLLGRELDAFVLYGSINGSGQA-----GYGANNGLDAL 3138  
OY 151 PRGAPEROLLPALACNG-----VRSLSLEADVRALGLACDLP-G 192

DB 3139 AR--YRRARGQAATVLMHMPMAGEGVTSELSQRLIRGVANMSPD-KALAGLEMAIRLG 3195  
OY 193 RFVASEAVLFLRL-----VSCPPGLDDQDEARALQGGPPYGPSTMSVTMDLR 247  
DB 3196 RTSVAIADVMSRFAPSFSAARPRP-LDGEIARRADESRGP---QPAAGCTALRDILL 3251  
OY 248 GLLPVLGPIIRISIPGIGVAAWRORSSRDPSWROPERTLRFRFRREKTAQPSGKKAR 307  
DB 3252 GLSEAEERRERVQLVASETA--VLGMTPSRLLDPDRGL----- 3289  
OY 308 EIDESLIFKKWELACVDAALLATQMDRVNAIPFTY-EOLDVLKHKIDELYPQGPESV 366  
DB 3290 --DLGLDSLMAVELSKRLQK--RTGMVPSLSPDHPTQSDVAMWLLLEQLTPQRPBPA 3344  
OY 367 IOHLGVLFKMSPEDIRKKN-----VTSLETKALLEVDKGHEMSPOAPR 411  
DB 3345 VR-----EVSREE--GWSPTIAIVGGLRMPGASDLESEWOVL-----VEERDTL 3388  
OY 412 RLP-----QVATLIDRFVGRGOL--DKDTLDTLTAFFPGYGLCSLSELSVPP----- 460  
DB 3389 RPIPAQREVEVALLYPDPDAKKTIVRNASLLDDVASFDPGF-GISREAEPMDPQRL 3447  
OY 461 --SSTWA-----VRPDLDTCDPRLDVLVYPKARLAFQNMNGSEYFVKIOSFLGAPTE 512  
DB 3448 LLETAMSALEDAGVAPRHLKGSDTG-----VFVGVAVSEY-----ASYRGSANE 3492  
OY 513 DLKALS 518  
DB 3493 DAYALT 3498

RESULT 10  
US-09-460-145-2  
Sequence 2, Application US/09460145  
Patent No. 6287838  
GENERAL INFORMATION:  
APPLICANT: Kriz, Ron  
APPLICANT: Song, Chuanzheng  
TITLE OF INVENTION: CYTOSOLIC PHOSPHOLIPASE A2-BETA ENZYMES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/460,145  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/788,975  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
REFERENCE/DOCKET NUMBER: G15289  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 778 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein



Db 579 WRPLAQAATHNRLGHHF--KDYFOHPHFSTW-----KATITD-----GLPNQLTPE 624  
QY 591 -----GYLVLDLSVOETLSGTFCPL 609  
Db 625 PHCLLDVGYLI-----NTSCL 641

## RESULT 12

US-08-827-208-3  
Sequence 3, Application US/08827208  
Patent No. 6025178  
GENERAL INFORMATION:  
APPLICANT: Chlou, Xue-Chiou C.  
APPLICANT: Kramer, Ruth M.  
APPLICANT: Pickard, Richard T.  
APPLICANT: Sharp, John D.  
APPLICANT: Striffler, Beth A.  
TITLE OF INVENTION: HUMAN PHOSPHOLIPASE A2 AND RELATED  
TITLE OF INVENTION: NUCLEIC ACID COMPOUNDS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,208  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/014,608  
FILING DATE: 29-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,264  
FILING DATE: 19-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Gaylo, Paul J.  
REGISTRATION NUMBER: 36,808  
REFERENCE/DOCKET NUMBER: X-10610  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317) 276-0756  
TELEFAX: (317) 276-3861  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-827-208-3

Query Match 3.5%; Score 113.5; DB 3; Length 913;  
Best Local Similarity 20.8%; Pred. No. 0.076;  
Matches 143; Conservative 72; Mismatches 257; Indels 215; Gaps 33;

QY 1 MALORLDPGMCGRPSGLILFLFSIGVHPARTLAGETGES--APLG-GVLTTPHNI 56  
Db 208 MELKVFDDDLVTGDP--VLSVLDAGLR-----AGFFRESSTLSQGGCRLEVERRL 260  
QY 57 SSLSPPRLIGFPACAEVSGSLSTERYRELAVALAOKNKLSTEQRLCLAHRLSEPPEDDAL 116  
Db 261 QSLADR-----GEMLVG---NGVLVARELSLHVQLLETGQKXSE 298  
QY 117 PLDLLLFLNPAPFSGPQACT-----RFF-SRITKANVDLIPGAPERQRLPALACWG 169

Db 299 HRVQLV--PGSCGEPQOEASVGTGTFRHCPACWEQELSLRLQDAPEEQALKAPLSALPSG 356  
QY 170 -----VRGSLLSADVR-----ALGLACDLPGRFAESAEVL---LPRVLS 208  
Db 357 OVRVLVPTSGEPPLMRVELKKEAGLRELAVRLEGEGPACAEQAFUSRRKQVAAALRLQLQ 416  
QY 209 CGPPLDQDOEAAARAALOGGCGPPYGPSTWSVSTMDALRGLLPVIGQ-----PIRSIPOG 264  
Db 417 LDGDLQDEDEIPVVAIMATGCG-----IRANTSLYGLAGLKEGLDLCVYITGA 466  
QY 265 IVAAWRORS-SRDPWRQPERTILRPRFREVEKTAQCPGSKKAREIDSLIFYKKMELEA 323  
Db 467 SGSTWALNIYEDPEWSQKD-----LAGPT----- 491  
QY 324 CYDALLATQMDRVNAIFFTYEQDLVLLKHKIDELYPQGYPPESVYIOHLYLFLKMSPEDIR 383  
Db 492 ---ELKTYQVYKKNLGLVADSQLQRYOEIAERARLDYPSCT----- 531  
QY 384 KWNYSLETLKALLEVDGKHEMSPOAP-----RRPLQVATLIDRFVKGRLDKDITDT 438  
Db 532 --NLWALLNEALLHDEPHDKHLSQGREALSHGQNPLIYCAL-----NTKGQ-----S 577  
QY 439 LTAFTPGYLCISPELSSVPPSSIMAVRPDDLTCD--PQDLVLYKARLAF-QANNK 495  
Db 578 LTTFEGGECFSPYEVGF--PKYGAFIPSELFEGSEFFMQLMKRLPESRICFLEGIMS 634  
QY 496 SEYFVKIOSFL--GGAPTE--DLKALSOQVNSMDLATFMKLR-----TDAVL 538  
Db 635 NLVANIQDLSLXMASPEPQOFMDRWVRNOANDKEVYPLKIEPSTGRIAEFTDLTT 694  
QY 539 --PLTVAEVQKILGPHVEGLKAERHRPVRWMLRQDDDLDTLGLGLOGGIPN----- 590  
Db 695 WRPLAQAATHNRLGHHF--KDYFOHPHFSTW-----KATITD-----GLPNQLTPE 740  
QY 591 -----GYLVLDLSVOETLSGTFCPL 609  
Db 741 PHCLLDVGYLI-----NTSCL 757

## RESULT 13

US-09-500-358-3  
Sequence 3, Application US/09500358  
Patent No. 6197569  
GENERAL INFORMATION:  
APPLICANT: Chlou, Xue-Chiou C.  
APPLICANT: Kramer, Ruth M.  
APPLICANT: Pickard, Richard T.  
APPLICANT: Sharp, John D.  
TITLE OF INVENTION: HUMAN PHOSPHOLIPASE A2 AND RELATED  
TITLE OF INVENTION: NUCLEIC ACID COMPOUNDS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/500,358  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,208  
FILING DATE: 28-MAR-1997  
APPLICATION NUMBER: US 60/014,608  
FILING DATE: 29-MAR-1996

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,264
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-500-358-3

```

```

Query Match          3.58; Score 113.5; DB 4; Length 913;
Best Local Similarity 20.88; Pred. No. 0.076; Indels 215; Gaps 33;
Matches 143; Conservative 72; Mismatches 257;

```

```

QY 1 MALORLPCWSCGDRPGSLFLFSLGWHVHPARTIAGETGES---APLG-GVLTTPHNI 56
DB 208 MELKVFODLVGTGDDP--VLSVLFDAQTLR-----AGEFRRESFSLSPQSGRLEVERRL 260
QY 57 SLSLPRQLLGFPCAEVSGSLSTERYRELAVALAOKNKLSTQRLCLAHRLSEPPEDDAL 116
DB 261 QSLADR-----GEMLYS---NGVLVARELSCLHYQLEETGDDKSSE 298
QY 117 PLDLLFLINPDAFSGPAQCT-----RFF-SRITKANVDLLPGAPERORLLPAAALCWG 169
DB 299 HRVOLLV--PQSCGPDQASVGTGTFRHCPACWQDELSTRLQDAPEQLKAPLSALPSG 356
QY 170 -----VRGSLSEADV-----ALGGLACDLPGRFYAESAEVL---LPRLVS 208
DB 357 QVRLVPTSOEPLMRVBLKKEAGRLAVRLGFGPCAEQAFLSRRKQVAAALRLQALQ 416
QY 209 CPGLDDQOEAAARALOGGPPYGPSTWSVTMDLRGLLPVLCQ---PIIRSIPOG 264
DB 417 LDGLODEIPVVAIMATGG-----IRAMTSLYGOLAGLKEGLLDVSYITGA 466
QY 265 IVAMRORS-SRQSWROPERTILRRPRREREVEKTACPSGKKAREIDSLIFYKKWLEEA 323
DB 467 SGTWLANLYEDPEWSQKD-----LACPT----- 491
QY 324 CVDALLATONDRVNAIPFTYEOLVLRKHLDELVPQGYPESVIOHLGYFLKMSPEDIR 383
DB 492 -----ELKTOYTKRKLGLVLAPOSQLORYRQELAEERARLGYPSCT----- 531
QY 384 KMWNTSLETKALLEVDKGMESPOAP-----RRPLPOVATLIRFVYKRGQLDKDILT 438
DB 532 --NLMALINEALHDEBDHKLSDQREALSHGNPLPYICAL-----NTKQO-----S 577
QY 439 LTAIYPGVGLCSPEELSSVPPSSIMAVRPQDLPTCD--PRQDLVLPKARLAF--QNWNG 495
DB 578 LITTEFEGWCFSEFYEVGF---PKYGAFLPSELGSEFFMQLMKRLPESNICELEGIWS 634
QY 496 SEFYVKIOSFL--GGAFTL--DLKALSOQVNSMDLATFMKLR-----TDAVL 538
DB 635 NLVYANLQDSLYMASSEPSQFMDRWVRNQNANDKQVPLKIEEPPSTAGRIAEFTDLIT 694
QY 539 --PLTVAEVOKLLDPHYEGLKAERHRVRWILRQRODDITDGLGLOGGIPN----- 590
DB 695 WRPLAQATHNLFRLGHLHN--KDYFOHPHFSTW---KATYLD-----GLPNQLPSE 740
QY 591 -----GVLYLDLSVQETLSGTPL 609
DB 741 PHCLLDVGYLI-----NTSCL 757

```

RESULT 14

```

US-09-498-809-3
; Sequence 3, Application US/09498809
; Patent No. 6242206
; GENERAL INFORMATION:
; APPLICANT: Chlou, Xue-Chiou C.
; APPLICANT: Kramer, Ruth M.
; APPLICANT: Pickard, Richard T.
; APPLICANT: Sharp, John D.
; APPLICANT: Stiffler, Beth A.
; TITLE OF INVENTION: HUMAN PHOSPHOLIPASE A2 AND RELATED
; TITLE OF INVENTION: NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/498,809
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,208
; FILING DATE:
; APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,264
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-498-809-3

```

```

Query Match          3.58; Score 113.5; DB 4; Length 913;
Best Local Similarity 20.88; Pred. No. 0.076; Indels 215; Gaps 33;
Matches 143; Conservative 72; Mismatches 257;

```

```

QY 1 MALORLPCWSCGDRPGSLFLFSLGWHVHPARTIAGETGES---APLG-GVLTTPHNI 56
DB 208 MELKVFODLVGTGDDP--VLSVLFDAQTLR-----AGEFRRESFSLSPQSGRLEVERRL 260
QY 57 SLSLPRQLLGFPCAEVSGSLSTERYRELAVALAOKNKLSTQRLCLAHRLSEPPEDDAL 116
DB 261 QSLADR-----GEMLYS---NGVLVARELSCLHYQLEETGDDKSSE 298
QY 117 PLDLLFLINPDAFSGPAQCT-----RFF-SRITKANVDLLPGAPERORLLPAAALCWG 169
DB 299 HRVOLLV--PQSCGPDQASVGTGTFRHCPACWQDELSTRLQDAPEQLKAPLSALPSG 356
QY 170 -----VRGSLSEADV-----ALGGLACDLPGRFYAESAEVL---LPRLVS 208
DB 357 QVRLVPTSOEPLMRVBLKKEAGRLAVRLGFGPCAEQAFLSRRKQVAAALRLQALQ 416
QY 209 CPGLDDQOEAAARALOGGPPYGPSTWSVTMDLRGLLPVLCQ---PIIRSIPOG 264
DB 417 LDGLODEIPVVAIMATGG-----IRAMTSLYGOLAGLKEGLLDVSYITGA 466

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:48:12 : Search time 143.5 Seconds  
(without alignments)  
1215.112 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALGRILDCWCSGDRPGSL.....ILGPPVLVLLASTLA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main: \*  
1: /cgn2\_6/ptodata/2/paa/PCPNUS.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/paa/US06.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/paa/US07.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/paa/US08.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/paa/US081.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/paa/US082.COMB.pep: \*  
7: /cgn2\_6/ptodata/2/paa/US083.COMB.pep: \*  
8: /cgn2\_6/ptodata/2/paa/US084.COMB.pep: \*  
9: /cgn2\_6/ptodata/2/paa/US085.COMB.pep: \*  
10: /cgn2\_6/ptodata/2/paa/US086.COMB.pep: \*  
11: /cgn2\_6/ptodata/2/paa/US087.COMB.pep: \*  
12: /cgn2\_6/ptodata/2/paa/US088.COMB.pep: \*  
13: /cgn2\_6/ptodata/2/paa/US089.COMB.pep: \*  
14: /cgn2\_6/ptodata/2/paa/US090.COMB.pep: \*  
15: /cgn2\_6/ptodata/2/paa/US091.COMB.pep: \*  
16: /cgn2\_6/ptodata/2/paa/US092.COMB.pep: \*  
17: /cgn2\_6/ptodata/2/paa/US093.COMB.pep: \*  
18: /cgn2\_6/ptodata/2/paa/US094.COMB.pep: \*  
19: /cgn2\_6/ptodata/2/paa/US095.COMB.pep: \*  
20: /cgn2\_6/ptodata/2/paa/US096.COMB.pep: \*  
21: /cgn2\_6/ptodata/2/paa/US097.COMB.pep: \*  
22: /cgn2\_6/ptodata/2/paa/US098.COMB.pep: \*  
23: /cgn2\_6/ptodata/2/paa/US099.COMB.pep: \*  
24: /cgn2\_6/ptodata/2/paa/US60.COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3055	93.7	622 8 US-08-426-819-35	Sequence 35, Appl
2	2956	90.6	584 8 US-08-426-819-36	Sequence 36, Appl
3	1547	47.4	328 19 US-09-513-597-2	Sequence 2, Appl
4	1547	47.4	402 19 US-09-513-597-13	Sequence 13, Appl
5	1531	46.9	399 19 US-09-513-597-1	Sequence 1, Appl
6	1247	38.2	248 8 US-08-426-819-37	Sequence 37, Appl
7	579.5	17.8	175 17 US-09-330-360-733	Sequence 733, App
8	559	17.1	144 17 US-09-330-360-599	Sequence 599, App
9	482	14.8	122 17 US-09-330-360-702	Sequence 702, App

10	424.5	13.0	134	21	US-09-758-458-406	Sequence 406, App
11	325	10.0	65	8	US-08-426-819-33	Sequence 33, Appl
12	220	6.7	144	21	US-09-758-458-404	Sequence 404, App
13	210	6.4	40	8	US-08-426-819-19	Sequence 19, Appl
14	163	5.0	35	8	US-08-426-819-5	Sequence 5, Appl
15	140	4.3	824	1	PCT-US01-14827-10981	Sequence 10981, A
16	132	4.0	2128	24	US-60-248-505-927	Sequence 927, App
17	131.5	4.0	2088	1	PCT-US01-14827-8424	Sequence 8424, App
18	128.5	3.9	89	22	US-09-875-043-532	Sequence 532, App
19	128.5	3.9	89	22	US-09-875-193-532	Sequence 532, App
20	125	3.8	2114	1	PCT-US01-08631-51615	Sequence 51615, A
21	125	3.8	2114	24	US-60-322-511-970	Sequence 970, App
22	124	3.8	962	21	US-09-758-455-449	Sequence 449, App
23	123.5	3.8	794	18	US-09-456-592-2	Sequence 2, Appl
24	123.5	3.8	794	21	US-09-789-404-2	Sequence 2, Appl
25	119	3.6	866	23	US-09-902-540-10786	Sequence 10786, A
26	119	3.6	943	21	US-09-738-626-5199	Sequence 5199, App
27	118	3.6	899	23	US-09-902-540-13735	Sequence 13735, A
28	117.5	3.6	3603	19	US-09-570-581A-1898	Sequence 1898, App
29	117.5	3.6	3603	19	US-09-573-655A-794	Sequence 794, App
30	117	3.6	1415	24	US-60-230-435-1076	Sequence 1076, App
31	117	3.6	1731	21	US-09-739-449-8331	Sequence 8331, App
32	117	3.6	1731	22	US-09-803-110-8331	Sequence 8331, App
33	116	3.6	1429	24	US-60-191-637-35218	Sequence 35218, A
34	116	3.6	1429	24	US-60-191-681-27637	Sequence 27637, A
35	115.5	3.5	842	21	US-09-577-304A-93	Sequence 93, Appl
36	115.5	3.5	842	19	US-09-758-282-93	Sequence 93, Appl
37	114.5	3.5	1016	24	US-60-208-099-22	Sequence 22, Appl
38	114.5	3.5	1016	24	US-60-208-587-35	Sequence 35, Appl
39	114.5	3.5	1310	24	US-60-212-413-205	Sequence 205, Appl
40	114	3.5	881	16	US-09-352-991A-31702	Sequence 31702, A
41	114	3.5	925	18	US-09-417-507-39237	Sequence 39237, A
42	114	3.5	970	24	US-60-242-679-871	Sequence 871, App
43	114	3.5	1067	24	US-60-161-932-2033	Sequence 2033, App
44	114	3.5	1067	24	US-60-167-217-13306	Sequence 13306, A
45	114	3.5	1067	24	US-60-173-464-10829	Sequence 10829, A

#### ALIGNMENTS

RESULT 1  
US-08-426-819-35  
: Sequence 35, Application US/08426819  
: GENERAL INFORMATION:  
: APPLICANT: Yamaguchi, Nozomi  
: APPLICANT: Kojima, Tetsuo  
: APPLICANT: Oh-Eda, Masayoshi  
: APPLICANT: Hattori, Kunhiro  
: TITLE OF INVENTION: Genes Coding for Megakaryocyte  
: NUMBER OF SEQUENCES: 37  
: CORRESPONDENCE ADDRESS:  
: ADDRESS: Birch, Stewart, Kolasch & Birch  
: STREET: P.O. Box 747  
: CITY: Falls Church  
: STATE: Virginia  
: COUNTRY: USA  
: ZIP: 22040-0747  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/426, 819  
: FILING DATE: 21-APR-1995  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Murphy Jr., Gerald M.  
: REGISTRATION NUMBER: 28,977  
: REFERENCE/DOCKET NUMBER: 230-107P  
: TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000  
 TELEFAX: 703-205-8050  
 INFORMATION FOR SEQ ID NO: 35:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 622 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-426-819-35

Query Match 93.7%; Score 3055; DB 8; Length 622;  
 Best Local Similarity 95.4%; Pred. No. 1e-275;  
 Matches 601; Conservative 3; Mismatches 16; Indels 10; Gaps 3;

```

QY 1 MALQRLDP-CMSGDRP-GSLLFLFSLGMYHPARTLAGETGTSAPSLGVTTPHNSS 58
DB 1 MALPAPRLDSCGPPALGSLFLFSLGMYOPSRITLAGTGEAPLDDGVLANPNISS 60
QY 59 LSPRLGFPCEAVSGSLSTERVELAVLAOKNKLSTEQRLCLAHRLSEPPEDLALPL 118
DB 61 LSPRLGFPCEAVSGSLSTERVELAVLAOKNKLSTEQRLCLAHRLSEPPEDLALPL 120
QY 119 DLLFLNPDAFSGPOACTRFRSRTTKANVDLLPRGAPERQRLPALACWGVRSLSSEA 178
DB 121 DLLFLNPDAFSGPOACTRFRSRTTKANVDLLPRGAPERQRLPALACWGVRSLSSEA 180
QY 179 DVALGGLACDPGRFVESAELLPRLVSCGPPDDQOEAAARALOGGPPYGPSTW 238
DB 181 DVALGGLACDPGRFVESAELLPRLVSCGPPDDQOEAAARALOGGPPYGPSTW 240
QY 239 SVSTMDALRGLLPVLOGPIIRSIPOGIYAAMRQSSRDPSWROPERTILRRFRREVET 298
DB 241 SVSTMDALRGLLPVLOGPIIRSIPOGIYAAMRQSSRDPSWROPERTILRRFRREVET 300
QY 299 ACPGSKKAREIDESLIFKKWELEACVDAALLATQMDRVNAIPFTYEQDLVKHKLDEL 358
DB 301 ACPGSKKAREIDESLIFKKWELEACVDAALLATQMDRVNAIPFTYEQDLVKHKLDEL 360
QY 359 POGYPEVIOHGLYFLKMSPEDIRKMNVTSLLETKALLEVDKHEMSPQAPRRRLPOVA 418
DB 361 POGYPEVIOHGLYFLKMSPEDIRKMNVTSLLETKALLEVDKHEMSPQAPRRRLPOVA 412
QY 419 TLIDRFVKGRLDKDLDLTLTAFYPGYGLCSLSPBELSSVPPSSITWAVRPDDLTCDPRQ 478
DB 413 TLIDRFVKGRLDKDLDLTLTAFYPGYGLCSLSPBELSSVPPSSITWAVRPDDLTCDPRQ 472
QY 479 LDVLYPKARLAFQNNNGSEYFVKIQSFLGAPTEDEKALSOQNNVSMDLATFNKLTDAVL 538
DB 473 LDVLYPKARLAFQNNNGSEYFVKIQSFLGAPTEDEKALSOQNNVSMDLATFNKLTDAVL 532
QY 539 PLTVAEVOKLGLPHEVGKAEERHPRVDMILRORODDLTGLGLOGGIPNGYVLDS 598
DB 533 PLTVAEVOKLGLPHEVGKAEERHPRVDMILRORODDLTGLGLOGGIPNGYVLDS 592
QY 599 VOETLSGTPCLLGGPVLTVALLASTLA 628
DB 593 VOETLSGTPCLLGGPVLTVALLASTLA 622

```

RESULT 2  
 US-08-426-819-36  
 Sequence 36, Application US/08426819  
 GENERAL INFORMATION:  
 APPLICANT: Yamaguchi, Nozomi  
 APPLICANT: Kojima, Tetsuo  
 APPLICANT: Oh-Eda, Masayoshi  
 APPLICANT: Hatori, Kunihiko  
 TITLE OF INVENTION: Genes Coding for Megakaryocyte  
 NUMBER OF SEQUENCES: 37  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747  
 CITY: Falls Church  
 STATE: Virginia  
 COUNTRY: USA  
 ZIP: 22040-0747  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/426,819  
 FILING DATE: 21-Apr-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murphy Jr., Gerald M.  
 REGISTRATION NUMBER: 28,977  
 REFERENCE/DOCKET NUMBER: 230-107P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-205-8000  
 TELEFAX: 703-205-8050  
 INFORMATION FOR SEQ ID NO: 36:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 584 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: NO  
 FRAGMENT TYPE: Internal  
 US-08-426-819-36

Query Match 90.6%; Score 2956; DB 8; Length 584;  
 Best Local Similarity 97.3%; Pred. No. 1.7e-266;  
 Matches 576; Conservative 2; Mismatches 6; Indels 8; Gaps 1;

```

QY 37 GETGTESAPLGGLVLTTPHNISLSLSPRLGFPCEAVSGSLSTERVELAVLAOKNKLST 96
DB 1 GETGEAPLDDGVLANPNISSLSLSPRLGFPCEAVSGSLSTERVELAVLAOKNKLST 60
QY 97 EQLRLARLSEPPEDLALPLDLLFLNPDAFSGPOACTRFRSRTTKANVDLLPRGAP 156
DB 61 EQLRLARLSEPPEDLALPLDLLFLNPDAFSGPOACTRFRSRTTKANVDLLPRGAP 120
QY 157 RORLPLALACWGVRSLSLEADYVALGGLACDLPGRFVESAELLPRLVSCGPPDDQ 216
DB 121 RORLPLALACWGVRSLSLEADYVALGGLACDLPGRFVESAELLPRLVSCGPPDDQ 180
QY 217 QOEAAARALOGGPPYGPSTWSVSTMDALRGLLPVLOGPIIRSIPOGIYAAMRQSSRD 276
DB 181 QOEAAARALOGGPPYGPSTWSVSTMDALRGLLPVLOGPIIRSIPOGIYAAMRQSSRD 240
QY 277 PSWRPERTILRRFRREVETACPSGKKAREIDESLIFKKWELEACVDAALLATQMDR 336
DB 241 PSWRPERTILRRFRREVETACPSGKKAREIDESLIFKKWELEACVDAALLATQMDR 300
QY 337 VNAIPFTYEQDLVKHKLDELTYPOGYPEVIOHGLYFLKMSPEDIRKMNVTSLLETKAL 396
DB 301 VNAIPFTYEQDLVKHKLDELTYPOGYPEVIOHGLYFLKMSPEDIRKMNVTSLLETKAL 360
QY 397 LEVDKHEMSPQAPRRRLPOVATLIDRFVKGRLDKDLDLTLTAFYPGYGLCSLSPBELS 456
DB 361 LEVDKHEMSPQAPRRRLPOVATLIDRFVKGRLDKDLDLTLTAFYPGYGLCSLSPBELS 412
QY 457 SVPPSSITWAVRPDDLTCDPRQDLVLYPKARLAFQNNNGSEYFVKIQSFLGAPTEDEKA 516
DB 413 SVPPSSITWAVRPDDLTCDPRQDLVLYPKARLAFQNNNGSEYFVKIQSFLGAPTEDEKA 472
QY 517 LSQNNVSMDLATFNKLTDAVLPLTVAEVOKLGLPHEVGKAEERHPRVDMILRORODD 576
DB 473 LSQNNVSMDLATFNKLTDAVLPLTVAEVOKLGLPHEVGKAEERHPRVDMILRORODD 532

```



QY 577 LDTGLGLOGGIPNGYLVLDLSVOETLSGTPCLLGGPVLTVALLASTLA 628  
|||||  
Db 533 LDTGLGLOGGIPNGYLVLDLSVOEALSGTPCLLGGPVLTVALLASTLA 584

RESULT 3  
US-09-513-597-2

; Sequence 2, Application US/09513597  
; GENERAL INFORMATION:  
; APPLICANT: Scholler, Nathalie B.  
; APPLICANT: Hellstrom, Ingegerd  
; APPLICANT: Hellstrom, Karl Erik  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: CARCINOMAS  
; FILE REFERENCE: 730033.410  
; CURRENT APPLICATION NUMBER: US/09/513.597  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-513-597-2

Query Match 47.4%: Score 1547; DB 19; Length 328;  
Best Local Similarity 97.1%: Pred. No. 3.6e-135;  
Matches 301; Conservative 1; Mismatches 0; Indels 8; Gaps 1;

QY 291 FRREVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVL 350  
|||||  
Db 1 FRREVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVL 60  
QY 351 KHKIDELYPGSPESVIOHGLYFLKMSPEDIRKMNWTSLETLKALLEVDKGHEKSPQAP 410  
|||||  
Db 61 KHKIDELYPGSPESVIOHGLYFLKMSPEDIRKMNWTSLETLKALLEVDKGHEKSPQAP 116  
QY 411 RRPDPVATLIDRFYKGGQDLDKDTLDTLTAFFPGYLCSSPEELSSVPSISIAVVRPQD 470  
|||||  
Db 117 ---PQVATLIDRFYKGGQDLDKDTLDTLTAFFPGYLCSSPEELSSVPSISIAVVRPQD 172  
QY 471 LDTCDPRQDVLTPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKALSOQVNSMDLATFM 530  
|||||  
Db 173 LDTCDPRQDVLTPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKALSOQVNSMDLATFM 232  
QY 531 KLRTDAVLPPLTVAEYQKLLGPHVEGLKAEERHRRVDMILRQRODDDLTGLGLOGGIPN 590  
|||||  
Db 233 KLRTDAVLPPLTVAEYQKLLGPHVEGLKAEERHRRVDMILRQRODDDLTGLGLOGGIPN 292  
QY 591 GYLVLDLSVQ 600  
|||||  
Db 293 GYLVLDLSVQ 302

RESULT 4  
US-09-513-597-13

; Sequence 13, Application US/09513597  
; GENERAL INFORMATION:  
; APPLICANT: Scholler, Nathalie B.  
; APPLICANT: Hellstrom, Ingegerd  
; APPLICANT: Hellstrom, Karl Erik  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: CARCINOMAS  
; FILE REFERENCE: 730033.410  
; CURRENT APPLICATION NUMBER: US/09/513.597  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 402  
; TYPE: PRT  
; ORGANISM: Homo sapien

US-09-513-597-13

Query Match 47.4%: Score 1547; DB 19; Length 402;  
Best Local Similarity 97.1%: Pred. No. 5.1e-135;  
Matches 301; Conservative 1; Mismatches 0; Indels 8; Gaps 1;

QY 291 FRREVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVL 350  
|||||  
Db 1 FRREVEKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVL 60  
QY 351 KHKIDELYPGSPESVIOHGLYFLKMSPEDIRKMNWTSLETLKALLEVDKGHEKSPQAP 410  
|||||  
Db 61 KHKIDELYPGSPESVIOHGLYFLKMSPEDIRKMNWTSLETLKALLEVDKGHEKSPQAP 116  
QY 411 RRPDPVATLIDRFYKGGQDLDKDTLDTLTAFFPGYLCSSPEELSSVPSISIAVVRPQD 470  
|||||  
Db 117 ---PQVATLIDRFYKGGQDLDKDTLDTLTAFFPGYLCSSPEELSSVPSISIAVVRPQD 172  
QY 471 LDTCDPRQDVLTPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKALSOQVNSMDLATFM 530  
|||||  
Db 173 LDTCDPRQDVLTPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKALSOQVNSMDLATFM 232  
QY 531 KLRTDAVLPPLTVAEYQKLLGPHVEGLKAEERHRRVDMILRQRODDDLTGLGLOGGIPN 590  
|||||  
Db 233 KLRTDAVLPPLTVAEYQKLLGPHVEGLKAEERHRRVDMILRQRODDDLTGLGLOGGIPN 292  
QY 591 GYLVLDLSVQ 600  
|||||  
Db 293 GYLVLDLSVQ 302

RESULT 5  
US-09-513-597-1

; Sequence 1, Application US/09513597  
; GENERAL INFORMATION:  
; APPLICANT: Scholler, Nathalie B.  
; APPLICANT: Hellstrom, Ingegerd  
; APPLICANT: Hellstrom, Karl Erik  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: CARCINOMAS  
; FILE REFERENCE: 730033.410  
; CURRENT APPLICATION NUMBER: US/09/513.597  
; CURRENT FILING DATE: 2000-02-25  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 399  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-513-597-1

Query Match 46.9%: Score 1531; DB 19; Length 399;  
Best Local Similarity 97.1%: Pred. No. 1.6e-133;  
Matches 298; Conservative 1; Mismatches 0; Indels 8; Gaps 1;

QY 294 EREKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVLKHK 353  
|||||  
Db 1 EREKTACPSGKKAREIDSLIFYKKWELEACVDALLATOMDRVNAIPFTYEQDVLKHK 60  
QY 354 LDELTPGSPESVIOHGLYFLKMSPEDIRKMNWTSLETLKALLEVDKGHEKSPQAP 413  
|||||  
Db 61 LDELTPGSPESVIOHGLYFLKMSPEDIRKMNWTSLETLKALLEVDKGHEKSPQAP 113  
QY 414 LPQVATLIDRFYKGGQDLDKDTLDTLTAFFPGYLCSSPEELSSVPSISIAVVRPQD 473  
|||||  
Db 114 -PQVATLIDRFYKGGQDLDKDTLDTLTAFFPGYLCSSPEELSSVPSISIAVVRPQD 172  
QY 474 CDPQDVLTPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKALSOQVNSMDLATFM 533  
|||||  
Db 173 CDPQDVLTPKARLAFQNMNGSEYFVKIQSFLGAPTEDEKALSOQVNSMDLATFM 232

QY 534 TDAVLPLTVAEYOKLLGPHVEGLKAEERHRRPVDMILRQRODDLTGLGLOGGIPNGYL 593  
DB 233 TDAVLPLTVAEYOKLLGPHVEGLKAEERHRRPVDMILRQRODDLTGLGLOGGIPNGYL 292  
QY 594 VLDLSVQ 600  
DB 293 VLDLSVQ 299

## RESULT 6

US-08-426-819-37  
; Sequence 37, Application US/08426819  
; GENERAL INFORMATION:  
; APPLICANT: Yamaguchi, Nozomi  
; APPLICANT: Kojima, Tetsuo  
; APPLICANT: Oh-Eda, Masayoshi  
; APPLICANT: Hattori, Kunihiko  
; TITLE OF INVENTION: Genes Coding for Megakaryocyte  
; TITLE OF INVENTION: Potentiator  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,819  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 230-107P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE: Internal  
; US-08-426-819-37

Query Match 38.2%; Score 1247; DB 8; Length 248;  
Best Local Similarity 97.6%; Pred. No. 2.6e-107;  
Matches 242; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 37 GETGESAPLGCVLTTPNINISLSRQLLGFPCAEGVSGLSTERYRELAVALAOKNVKLSL 96  
DB 1 GETGEAPALGCVLTTPNINISLSRQLLGFPCAEGVSGLSTERYRELAVALAOKNVKLSL 60  
QY 97 EOLRLARLSEPPEDDALPLDILLFLNPDAGSGPOACTRFRSRTKANVDLPRGAPAE 156  
DB 61 EOLRLARLSEPPEDDALPLDILLFLNPDAGSGPOACTRFRSRTKANVDLPRGAPAE 120  
QY 157 RORLPAALACMGVAGSLSEADVAGLACDLPRFVAESAENVLPRLVSCPPPLDD 216  
DB 121 RORLPAALACMGVAGSLSEADVAGLACDLPRFVAESAENVLPRLVSCPPPLDD 180  
QY 217 OOEARAALOGGPGPPSTWSTMDALRGLPLVLCOPITIRISPGIVAAAMRORSSRD 276

DB 181 OOEARAALOGGPGPPSTWSTMDALRGLPLVLCOPITIRISPGIVAAAMRORSSRD 240  
QY 277 PSMROPER 284  
DB 241 PSMROPER 248

## RESULT 7

US-09-330-360-733  
; Sequence 733, Application US/09330360  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Robison, Keith E.  
; APPLICANT: Holtzman, Douglas A.  
; TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
; TITLE OF INVENTION: Human Fetal Lung Library  
; FILE REFERENCE: MLN98-34PA  
; CURRENT APPLICATION NUMBER: US/09/330,360  
; CURRENT FILING DATE: 1999-06-11  
; PRIOR APPLICATION NUMBER: 60/090,258  
; PRIOR FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 1102  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 733  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(33)  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(175)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
; US-09-330-360-733

Query Match 17.8%; Score 579.5; DB 17; Length 175;  
Best Local Similarity 68.6%; Pred. No. 3.8e-45;  
Matches 129; Conservative 5; Mismatches 23; Indels 31; Gaps 5;

QY 1 MALORLDP-CWCSGRP-GSLIFLLFSIGWHPARTLAGGCTESAPLGVTTTPHNIS 58  
DB 1 MALPARPLGSCGTPALGSLLEFLSLGWOPSTTLAGEGQEAAPLDGVLANPNISS 60  
QY 59 LSPROLGFPCEAVSGLSTERYRELAVALAOKNVKLSLEOLRCLAHRLSEPPEDDALPL 118  
DB 61 LSPROLGFPCEAVSGLSTERYRELAVALAOKNVKLSLEOLRCLAHRLSEPPEDDALPL 120  
QY 119 DLLFLNPDAGSGPOACTRFRSRTKANVDLPRGAPERORLL-----PALAC- 167  
DB 121 DLLFLNPDAGSGPOA-----XPVSSPASRRPMWTCSESGSRATAAAG 165  
QY 168 ---WGVRG 172  
DB 166 SGLMGVVG 173

## RESULT 8

US-09-330-360-599  
; Sequence 599, Application US/09330360  
; GENERAL INFORMATION:  
; APPLICANT: Gearing, David P.  
; APPLICANT: Robison, Keith E.  
; APPLICANT: Holtzman, Douglas A.  
; TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
; TITLE OF INVENTION: Human Fetal Lung Library  
; FILE REFERENCE: MLN98-34PA  
; CURRENT APPLICATION NUMBER: US/09/330,360  
; CURRENT FILING DATE: 1999-06-11  
; PRIOR APPLICATION NUMBER: 60/090,258  
; PRIOR FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 1102  
; SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 599  
LENGTH: 144  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(33)  
NAME/KEY: VARIANT  
LOCATION: (1)...(144)  
OTHER INFORMATION: Xaa - Any Amino Acid  
US-09-330-360-599

Query Match 17.1%; Score 559; DB 17; Length 144;  
Best Local Similarity 86.0%; Pred. No. 2.2e-43;  
Matches 117; Conservative 2; Mismatches 15; Indels 2; Gaps 2;

QY 1 MALORLDP-CMSCGDRP-GSLFLFLFSLGWHPARTLAGETTESAPLGVLTTPHNIS 58  
DB 1 MALPTARLFGSCGTPALGSLFLFLFSLGWQPSRTLAGETGEAPLDGVLANPNIS 60  
QY 59 LSPRLGLFPCEAVSGSLTERVRELAVLAQKNVLSLEQLCLAHRLSEPPEDIDALPL 118  
DB 61 LSPRLGLFPCEAVSGSLTERVRELAVLAQKNVLSLEQLCLAHRLSEPPEDIDALPL 120  
QY 119 DULFLNPDATFSGPOA 134  
DB 121 DULFLNPDATFSGPOA 136

RESULT 9  
US-09-330-360-702  
Sequence 702, Application US/09330360  
GENERAL INFORMATION:  
APPLICANT: Gearing, David P.  
APPLICANT: Robison, Keith E.  
APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
FILE REFERENCE: M1N98-34PA  
CURRENT APPLICATION NUMBER: US/09/330,360  
CURRENT FILING DATE: 1999-06-11  
PRIOR APPLICATION NUMBER: 60/090,258  
PRIOR FILING DATE: 1998-06-22  
NUMBER OF SEQ ID NOS: 1102  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 702  
LENGTH: 122  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(33)  
US-09-330-360-702

Query Match 14.8%; Score 482; DB 17; Length 122;  
Best Local Similarity 84.3%; Pred. No. 2.7e-36;  
Matches 102; Conservative 2; Mismatches 15; Indels 2; Gaps 2;

QY 1 MALORLDP-CMSCGDRP-GSLFLFLFSLGWHPARTLAGETTESAPLGVLTTPHNIS 58  
DB 1 MALPTARLFGSCGTPALGSLFLFLFSLGWQPSRTLAGETGEAPLDGVLANPNIS 60  
QY 59 LSPRLGLFPCEAVSGSLTERVRELAVLAQKNVLSLEQLCLAHRLSEPPEDIDALPL 118  
DB 61 LSPRLGLFPCEAVSGSLTERVRELAVLAQKNVLSLEQLCLAHRLSEPPEDIDALPL 120  
QY 119 D 119  
DB 121 D 121

RESULT 10  
US-09-758-458-406  
Sequence 406, Application US/09758458  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM041  
CURRENT APPLICATION NUMBER: US/09/758,458  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 520  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 406  
LENGTH: 134  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (51)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (56)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (97)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (117)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (127)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (131)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-758-458-406

Query Match 13.0%; Score 424.5; DB 21; Length 134;  
Best Local Similarity 63.7%; Pred. No. 7.5e-31;  
Matches 93; Conservative 3; Mismatches 23; Indels 27; Gaps 3;

QY 263 QGIVAAHROSRSDPSWRPRTILRRPRREVEKTACSGKKARIDSLFYKKMELE 322  
DB 7 RSGGKWR-----RQPVLODEGPRDRREPH-----LQEVGAG----- 39  
QY 323 ACVDALLATOMDRVNAIPFTYEOLDVLRKHKLDELXPQGPSPVIOHGLFLKNSPEDI 382  
DB 40 TCVDALLATQXDRVNAIPFTYEOLDVLRKHKLDELXPQGPSPVIOHGLFLKNSPEDI 99  
QY 383 RKWNTSLETLKALLEVDKGHMSPO 408  
DB 100 RKWNTSLETLKALLEVDKGHMSPO 125

RESULT 11  
US-08-426-819-33  
Sequence 33, Application US/08426819  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, Nozomi  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiko  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church

STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 65 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-426-819-33

Query Match 10.0%; Score 325; DB 8; Length 65;  
Best Local Similarity 100.0%; Pred. No. 4,5e-22;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 LIGFCAEVSGLSTERVRELAVALAKNKVLTSEOLRCIAHRLSEPPEDLALPLDLILF 123  
DB 1 LIGFCAEVSGLSTERVRELAVALAKNKVLTSEOLRCIAHRLSEPPEDLALPLDLILF 60  
QY 124 LNPDA 128  
DB 61 LNPDA 65

RESULT 12  
US-09-758-458-404  
Sequence 404, Application US/09758458  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM041  
CURRENT APPLICATION NUMBER: US/09/758,458  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 520  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 404  
LENGTH: 144  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (128)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (137)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (140)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-758-458-404

Query Match 6.7%; Score 220; DB 21; Length 144;  
Best Local Similarity 47.5%; Pred No. 1.1e-11;  
Matches 58; Conservative 3; Mismatches 11; Indels 50; Gaps 5;

QY 4 QRLDPCW-----SCG-----DRPSLLFLFLSLGWHPARTLGE 38  
DB 25 QRLDPCWGVGPPPSAASCSSASDGCSPRGWLERQGRRL-----HPWTDLA-- 73  
QY 39 TGTESAPLGAVLTPTNINISLSPROLLGFPCAEVSGLSTERVRELAVALAKNKV-LSTE 97  
DB 74 -----NPPNISLSROLLGFPCAEVSGLSTERVRELAVALAKNKVLAQPE 120  
QY 98 QL 99  
DB 121 QI 122

RESULT 13  
US-08-426-819-19  
Sequence 19, Application US/08426819  
GENERAL INFORMATION:  
APPLICANT: Yamaguchi, Nozomi  
APPLICANT: Kojima, Tetsuo  
APPLICANT: Oh-Eda, Masayoshi  
APPLICANT: Hattori, Kunihiko  
TITLE OF INVENTION: Genes Coding for Megakaryocyte  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,819  
FILING DATE: 21-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 230-107P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 amino acids  
TYPE: amino acid  
SPRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: Internal  
IMMEDIATE SOURCE:  
CLONE: HPCY5  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..40  
OTHER INFORMATION: /label= fragment  
OTHER INFORMATION: /note="sequence of cyanogen bromide fragment of MegPOT fr  
OTHER INFORMATION: CDNA in Table 3"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 32  
OTHER INFORMATION: /product="1s Xaa in COS"  
FEATURE:

```
NAME/KEY: Modified-site
LOCATION: 35
OTHER INFORMATION: /product= "is Xaa in COS"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 37
OTHER INFORMATION: /product= "is Xaa in COS"
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
OTHER INFORMATION: /label= fragment
OTHER INFORMATION: /note= "sequence of cyanogen bromide fragment representing
OTHER INFORMATION: C-terminus of MegPOT determined in Example 28"
US-08-426-819-19

Query Match
Best Local Similarity 100.0%; Score 210; DB 8; Length 40;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 244 DALRGLPVIGPPIIRSIPOGIVAAWRORSRDPSPWROPE 283
DB 1 DALRGLPVIGPPIIRSIPOGIVAAWRORSRDPSPWROPE 40

RESULT 14
US-08-426-819-5
Sequence 5, Application US/08426819
GENERAL INFORMATION:
APPLICANT: Yamaguchi, Nozomi
APPLICANT: Kojima, Tetsuo
APPLICANT: Oh-Eda, Masayoshi
APPLICANT: Hatcori, Kunihiko
TITLE OF INVENTION: Genes Coding for Megakaryocyte
TITLE OF INVENTION: Potentiator
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,819
FILING DATE: 21-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 230-107P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 35 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHEtical: NO
FRAGMENT TYPE: Internal
IMMEDIATE SOURCE:
CLONE: HPCY5
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..35
```

```
OTHER INFORMATION: /label= peptide
OTHER INFORMATION: /note= "sequence of a Glu-C peptide of MegPOT"
FEATURE:
NAME/KEY: Peptide
LOCATION: 6..33
OTHER INFORMATION: /label= peptide
OTHER INFORMATION: /note= "region of Glu-C peptide having low frequency of us
OTHER INFORMATION: codons."
US-08-426-819-5

Query Match
Best Local Similarity 97.1%; Score 163; DB 8; Length 35;
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 83 LAVALAOKNVKLTSTQRLAHLRSEPEDDALP 117
DB 1 LAVALAOKNVKLTSTQRLAHLRSEPEDDALP 35

RESULT 15
PCT-US01-14827-10981
Sequence 10981, Application PC/TUS0114827
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-104
CURRENT APPLICATION NUMBER: PCT/US01/14827
CURRENT FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: 09/577,408
PRIOR FILING DATE: 2000-05-18
NUMBER OF SEQ ID NOS: 16102
SOFTWARE: Custom
SEQ ID NO 10981
LENGTH: 824
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (747)..(768)
OTHER INFORMATION: POLYCYSTIC KIDNEY DISEASE PROTEIN SIGNATURE domain identified
OTHER INFORMATION: by EMATRIX, accession number PR00500B, p-value=4.316e-09, raw
PCT-US01-14827-10981

Query Match
Best Local Similarity 23.3%; Score 140; DB 1; Length 824;
Matches 155; Conservative 70; Mismatches 267; Indels 172; Gaps 32;

OY 15 RPSGL--LFLFSLGWHPARTLAGETGESAPLGVLTPPHN-ISSISPQLLGFPCAE 71
DB 194 RPTFGALGALATLNLNHNALVYLRPMAFQGLLRWRLSLHNLVSLVAPRALGLPALR 253
OY 72 VSGLSTERELAVALAOKNVKLTSTQRLAHLRSEPP-----EDLDALP--LDLL 122
DB 254 RLSLHNHNLQALPGV-----LSQARGLARLELGHNPITYAGEEDGLALPGLRELL 305
OY 123 -----FLNPAFSGPQACTRFESRITKAN--VDLLP-----RGAPRQRLLAALAC----- 167
DB 306 DGGALQALGPRAFA--HCPRLHTLDLRGNOLDLTPPLQGGQLRRLRLQGNPLMCGQA 362
OY 168 -----WGVRSLSLEADVRALGLACDLPGRFVASEVLLPRLVSCPGPLDQDQEAR 222
DB 363 RPLEW-----LARKVRSNG--ACGPRRLRGELDALRPMDLRCGDAAQEEELIE 414
OY 223 AALQ-----GPPYGPSTWSVSTMDALRGLPLVIGPI-----INSIPOGIVAAWR 272
DB 415 RAVAGPRAPRPGPGPEERAVAPCPRACVCPESHSCEGGLQAVPRGF----- 467
OY 273 SSRPSMRQPERTILRPFRREVEKTCPSGKKAREIDSLIFKKWLEACVDAALLAT 332
DB 468 ---PS--DTQLDLRRNHFPSPRAAP-----GLGHVLSLHLQHGIALELAG 511
```

```

OY 333 QMDRVNAIPFTY---EOLDVLKHKLDELYPQGYPESVIOHGLYFLKMSPEDIRKWNVTS 389
Db 512 ALAGLGRLIYLYLSDNOLAGLSAALLEGAPR-----LGYLYERN----- 551
OY 390 LETTKALLEVDKGHMSPOAPRRPLQVATLLIDRFVKGRGOLDKDLDTLTAFYPGYL-- 447
Db 552 ---RFLQV-----PGAALRALPSLSL-----HLQDNADVRLA---PGDLGR 587
OY 448 -----CSLSPSELSSVPSSIMAVRPQDLDTCPRLQDLVLYPKARLAFQNMNGSEYFV 500
Db 588 TRALRMGHLHSGNRITFVSLGALGPARELEKLIHLDNRQDLREVPTGA-----LEGIPALL 641
OY 501 KIOSEFGAPTEDELKALSQQWNSMDLATEFMKLRDVAVLPLTVAEVOKLIGPHEGL--KA 558
Db 642 ELQ--LSGNPLRALPDGAFQPYGRSLQH-----LFLNSSGLEQYGTGHLAGLYQEA 690
OY 559 EERRRPVRDWILRQRODLDLTGLGLGGIPNGYLVLDLSVOETLSGTPCLLGGPVLTV 618
Db 691 AQGH-----RQRAFVQQAFAASPYPVPGIGPGLQSLHLQ-KNQLRALPAL-----PSLSQ 737
OY 619 LALL 622
Db 738 LELI 741

```

Search completed: December 7, 2001, 09:52:20  
 Job time: 248 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:48:02 ; Search time 26.14 Seconds  
(without alignments) 244.470 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALQRIDPCWCGDRPGSL.....LIGPFLVTLALLASTLA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 57046 seqs, 10175881 residues

Total number of hits satisfying chosen parameters: 57046

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	579.5	17.8	175	US-09-828-792-733	Sequence 733, App
2	559	17.1	144	US-09-828-792-599	Sequence 599, App
3	482	14.8	122	US-09-828-792-702	Sequence 702, App
4	131.5	4.0	1159	US-09-815-242-11089	Sequence 2, Appli
5	102	3.1	1159	US-09-815-242-11089	Sequence 11089, A
6	101.5	3.1	2472	US-09-815-242-5064	Sequence 5064, Ap
7	100	3.1	956	US-09-815-242-11925	Sequence 11925, A
8	99.5	3.1	941	US-09-722-487-55	Sequence 55, Appli
9	99.5	3.1	941	US-09-722-708-55	Sequence 55, Appli
10	99.5	3.1	3739	US-09-657-440-2	Sequence 2, Appli
11	99	3.0	610	US-09-325-795-4	Sequence 4, Appli
12	96.5	3.0	308	US-09-976-594-449	Sequence 188, App
13	96.5	3.0	308	US-09-981-353-188	Sequence 629, Ap
14	95.5	2.9	689	US-09-897-516-6629	Sequence 5107, Ap
15	95.5	2.9	4342	US-09-815-242-5107	Sequence 5111, Ap
16	95	2.9	1062	US-09-815-242-5111	Sequence 10062, A
17	94.5	2.9	1048	US-09-815-242-10062	Sequence 12811, A
18	93	2.9	667	US-09-815-242-12811	Sequence 31, Appli
19	93	2.9	3816	US-09-808-880-3	Sequence 51, Appli
20	92.5	2.8	937	US-09-964-899-51	Sequence 8074, Ap
21	92	2.8	353	US-09-897-516-8074	Sequence 5408, Ap
22	92	2.8	657	US-09-815-242-5408	Sequence 12490, A
23	92	2.8	667	US-09-815-242-12490	Sequence 4688, Ap
24	92	2.8	1527	US-09-897-516-4688	Sequence 6555, Ap
25	91.5	2.8	1656	US-09-897-516-6555	Sequence 2, Appli
26	91	2.8	349	US-09-919-935-2	Sequence 31, Appli
27	90	2.8	451	US-09-964-899-31	

28	90	2.8	519	US-09-897-516-6004	Sequence 6004, Ap
29	89.5	2.7	9799	US-09-965-728-146	Sequence 146, App
30	89.5	2.7	11721	US-09-965-728-162	Sequence 162, App
31	89	2.7	1115	US-09-915-543-17	Sequence 17, Appli
32	88.5	2.7	200	US-09-815-242-5053	Sequence 5053, Ap
33	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
34	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
35	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
36	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
37	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
38	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
39	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
40	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
41	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
42	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
43	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
44	88.5	2.7	775	US-09-989-723-326	Sequence 326, App
45	88.5	2.7	775	US-09-989-723-326	Sequence 326, App

ALIGNMENTS

RESULT 1  
US-09-828-792-733  
Sequence 733, Application US/09828792  
GENERAL INFORMATION:  
APPLICANT: Georing, David P.  
APPLICANT: Robison, Keith E.  
APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: Nucleic Acid Molecules Derived from a  
FILE REFERENCE: MLN98-349a  
CURRENT APPLICATION NUMBER: US/09/828,792  
CURRENT FILING DATE: 2001-04-09  
PRIOR APPLICATION NUMBER: US/09/330,360  
PRIOR FILING DATE: 1999-06-11  
PRIOR APPLICATION NUMBER: 60/090,258  
PRIOR FILING DATE: 1998-06-22  
NUMBER OF SEQ ID NOS: 1102  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 733  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(33)  
NAME/KEY: VARIANT  
LOCATION: (1)...(175)  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-828-792-733

Query Match 17.8%; Score 579.5; DB 5; Length 175;  
Best Local Similarity 68.6%; Pred. No. 2.4e-42;  
Matches 129; Conservative 5; Mismatches 23; Indels 31; Gaps 5;  
QY 1 MALQRIDPCWCGDRPGSLFLFLFSIGWVHPARTLAGETESAPLGGVLTTPHNSS 58  
DB 1 MALPAPRLSCGCPALGSLFLFLFSIGWVQPSSTLAGEGAEAPLDVLANPNISS 60  
QY 59 LSPROLLGFCFAEYSGSLSTERYRELAVLAQKNVLSLEQLRCLAHRLSEPPEDDALPL 118  
DB 61 LSPROLLGFCFAEYSGSLSTERYRELAVLAQKNVLSLEQLRCLAHRLSEPPEDDALPL 120  
QY 119 DLLFLNDAFSGQACTRFPSTRITKANVDDLPRCAPERQRL-----PALAAC- 167  
DB 121 DLLFLNDAFSGQACTRFPSTRITKANVDDLPRCAPERQRL-----PALAAC- 165  
QY 168 ---MGVRC 172  
DB 166 SGLMGVRC 173

```
RESULT 2
US-09-828-792-599
; Sequence 599, Application US/09828792
; GENERAL INFORMATION:
; APPLICANT: Gearling, David P.
; APPLICANT: Robison, Keith E.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: Nucleic Acid Molecules Derived from a
; TITLE OF INVENTION: Human Fetal Lung Library
; FILE REFERENCE: MLN98-34Pa
; CURRENT APPLICATION NUMBER: US/09/828,792
; CURRENT FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US/09/330,360
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/090,258
; PRIOR FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 1102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 599
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(33)
; NAME/KEY: VARIANT
; LOCATION: (1)...(144)
; OTHER INFORMATION: Xaa - Any Amino Acid
US-09-828-792-599
```

```
Query Match 17.1%; Score 559; DB 5; Length 144;
Best Local Similarity 86.0%; Pred. No. 9.9e-41;
Matches 117; Conservative 2; Mismatches 15; Indels 2; Gaps 2;

QY 1 MALQRLDP-CMSCGRP-GSLIFLFSIGVHPARTLAGETGESAPLGGVITTPHNIS 58
Db 1 MALPTARPLGSCGTPALGSLIFLFSIGVQPSRTLAGETGQEAAPLDGVLANPNISS 60

QY 59 LSPRLGFPCEAVSGLSTERYRELAVALAOKNVKLTSTEQRLCLAHRLSEPPEDDALPL 118
Db 61 LSPRLGFPCEAVSGLSTERYRELAVALAOKNVKLTSTEQRLCLAHRLSEPPEDDALPL 120

QY 119 DILFLNPDAFSGPQA 134
Db 121 DILFLNPDAFSGPQA 136
```

```
RESULT 3
US-09-828-792-702
; Sequence 702, Application US/09828792
; GENERAL INFORMATION:
; APPLICANT: Gearling, David P.
; APPLICANT: Robison, Keith E.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: Nucleic Acid Molecules Derived from a
; TITLE OF INVENTION: Human Fetal Lung Library
; FILE REFERENCE: MLN98-34Pa
; CURRENT APPLICATION NUMBER: US/09/828,792
; CURRENT FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US/09/330,360
; PRIOR FILING DATE: 1999-06-11
; PRIOR APPLICATION NUMBER: 60/090,258
; PRIOR FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 1102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 702
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: SIGNAL
; LOCATION: (1)...(33)
US-09-828-792-702
```

```
Query Match 14.8%; Score 482; DB 5; Length 122;
Best Local Similarity 84.3%; Pred. No. 2.8e-34;
Matches 102; Conservative 2; Mismatches 15; Indels 2; Gaps 2;

QY 1 MALQRLDP-CMSCGRP-GSLIFLFSIGVHPARTLAGETGESAPLGGVITTPHNIS 58
Db 1 MALPTARPLGSCGTPALGSLIFLFSIGVQPSRTLAGETGQEAAPLDGVLANPNISS 60

QY 59 LSPRLGFPCEAVSGLSTERYRELAVALAOKNVKLTSTEQRLCLAHRLSEPPEDDALPL 118
Db 61 LSPRLGFPCEAVSGLSTERYRELAVALAOKNVKLTSTEQRLCLAHRLSEPPEDDALPL 120

QY 119 D 119
Db 121 D 121
```

```
RESULT 4
US-60-325-795-2
; Sequence 2, Application US/60325795
; GENERAL INFORMATION:
; APPLICANT: Su, Eric W
; APPLICANT: Wang, He
; APPLICANT: Zhi, Yu
; TITLE OF INVENTION: NOVEL SECRETED PROTEINS AND THEIR USES
; FILE REFERENCE: P-14990
; CURRENT APPLICATION NUMBER: US/60/325,795
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-325-795-2
```

```
Query Match 4.0%; Score 131.5; DB 6; Length 762;
Best Local Similarity 24.4%; Pred. No. 0.0028;
Matches 113; Conservative 45; Mismatches 181; Indels 125; Gaps 23;
```

```
QY 15 RRGSL--LEFLFSIGVHPARTLAGETGESAPLGGVITTPHN--ISSLSPRLGFPCE 71
Db 150 RRGTFGALGALATINLAHNAVLYCPMAFQGLLRVRLSHNALSVLAPALGLPALR 209

QY 72 VSGLSTERYRELAVALAOKNVKLTSTEQRLCLAH--RLSEPP-----EDLDALP--LDL 122
Db 210 RLSLHNLQALPGPV-----LSQARGLARLELGNPLTYAGEEGLALPGIRELL 261

QY 123 -----FLNPDAFSGPQACTRFFSRITKAN--VDLLP-----RGABERQRLPALAAC----- 167
Db 262 DGALQALGPRPFA--HCPRLHTLDLNGNDLTPPLQGGGQLRRLLQGNPLMCGQA 318

QY 168 -----WGVKSLSEADYRALGGLACDLPGRFVAESAVALPRLVYSCGCPDLDDQOEAR 222
Db 319 RPLLEW-----LARAARVSDG--ACQPRRLRGEALDLPMDLRCPGDAAGEEEL 370

QY 223 AALQG-----GGPRYGPSTWSYSTMDALRGELPLVVGPI-----IRSIPOGIYAAMROR 272
Db 371 RAVAGPRAPRPGPRGPEERAVAPCPRAVCVPSRHSSCEGGLQAVPGRF----- 423

QY 273 SSRPSPWQRPRTILRPFRREVEKTACPSGKKAREIDESTLFFKKWLEACVDAALLAT 332
Db 424 ----PS--DTQLDLRRHHPSPVRAAP-----GLGHVLSHLQHCIGTALEAG 467

QY 333 QMDRYNAIPFTY---EQLDVAKHKLDLDPGYPSVYIQLGYFLKMSPEIDRKWNTS 389
Db 468 ALAGIGRLIYILSDNQLAGLSAALLEGAPR-----LGYLYLERN----- 507
```



```
OY 390 LETLKALLEVDKHEMSPQAPRRPLPOVATL-----IDREFVG 427
Db 508 -----RFLQV-----FGAALRALPSLFLHLDONNAVDR LAPC 539

RESULT 5
US-09-815-242-11089
: Sequence 11089, Application US/09815242
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA 011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11089
: LENGTH: 1159
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
US-09-815-242-11089

Query Match 3.1%; Score 102; DB 5; Length 1159;
Best Local Similarity 20.2%; Pred. No. 1.7;
Matches 149; Conservative 96; Mismatches 272; Indels 222; Gaps 34;

OY 29 VHPARTLAGEGTESAPLPGVLTTPHNISSLSR-----QLLGFPCA 70
Db 495 INMARKLEGVTRMKGNAAGVVISPTLITDFAPLYCDNEGLHPVTHPKNDVEYAGLYKF 554
OY 71 EVSGLSREVELAVALLAOKNVKLTSTPOLRCIAHRLSEPPELDALPLDLLFLNPDASF 130
Db 555 DFLGLRTLTIIKMLDIT--NVBMVRE-----GKPRVDIAIPLD-----DPEEFE 598
OY 131 --GPOACTREFSRITKANVDLLPRGAPRORLLPALACMGVRSLSLEADYALGLAC 188
Db 599 LLKRSETTAVFOLSESGMKDLIKRLQDP-----CF-----EDIAL--VAL 637
OY 189 DLPRGFVAESAENVLLPRL-----VSCPRPLDODQDEA--RAALOGGGPPYPRPSWST 242
Db 638 FRPGFLOGVWDFNLDNRHGREEVSY--DAEYOHASLKLPLE--PTYG-----682
OY 243 MDALRGLLPVLOPIIRISIPQIYAAV-----RORSSRDPSPMRQPERTILPRFR 293
Db 683 -----IILXQEVVMQIAQ-VLAGTYLGGADLLRRAMGKKKPPEMAKORIVREGEAK 733
OY 294 E-----VEKTACPSGKKAREIDSLIFYK---KWELEACVDALLATQMDR 336
Db 734 NGIDGELSMKIFDLVEKFPAGYGFMKSHSAAYALVSYQTLMLKTHRPABFMAAVMTSEMDN 793
```

```
OY 337 VNAIPETVEQ-----LDVLKHLDELVPOG---YPESVIOHGX-----372
Db 794 TDKIVGLYDECLRMGLKATVPDINIGKHNF--SVNEQGIYVIGAIKVGCGPIEALVAA 852
OY 373 -----LFLKMSPEDIRKMNVTSETLKALLEVDKHEMSPQ--APRRPLPOVATLI 421
Db 853 RNEGIGFKDLFDLCARVLDKIKINRTFESLMSGAFDK--LGPFRALSLNLEDAALRAS 909
OY 422 DREYKRGQDLKDTLDTL-----AFYPGYLC--LSPE-----LSSVPPSI 463
Db 910 DOHAKDEAMGQDMEGVLTETHEDVENAVANTPYTEKOIILDGERETGLYLSHPVSRY 969
OY 464 WAVRPODLDTCPDOLVLYPKARLAFQNMNGSEFVFIQ-----SFLGAPTEDLKALS 518
Db 970 L-----KELSHYSTRKLKLANRRGQISTVAGLYVASIANTKGNRLGATLDD-----1020
OY 519 QONVSMDLATF-----MKLRTDVAV-----LPLTVAEVOKL--LGP 552
Db 1021 -RSGRDLTLFGESLQGEKGLQKDTVIVASQVSDFFSGGLKMTVRELMTLDEARSR 1079
OY 553 VEGKAEERHPRVDMILRQKDDLD-----TLGLGLOGGIPNGYVLVDLSVOETLSGP 607
Db 1080 VKSLAISLSEHQITPSPFKOPKALPEVSGGLPIINVYQSPFKRGLRLGLQVMSIIP 1138
OY 608 CLLGPGVLTVALLLAST 626
Db 1139 -----DELTELVLNLSGS 1152

RESULT 6
US-09-815-242-5064
: Sequence 5064, Application US/09815242
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA 011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5064
: LENGTH: 2472
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5064

Query Match 3.1%; Score 101.5; DB 5; Length 2472;
Best Local Similarity 21.7%; Pred. No. 5.6;
Matches 138; Conservative 70; Mismatches 216; Indels 211; Gaps 33;
```



ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/722.487  
FILING DATE: 28-Nov-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/179,558  
FILING DATE: <Unknown>  
APPLICATION NUMBER: U.S. 60/063,898  
FILING DATE: 31-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 9426-005-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 941 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 55:  
US-09-722-487-55

Query Match 3.1%; Score 99.5; DB 5; Length 941;  
Best Local Similarity 20.8%; Pred. No. 2;  
Matches 148; Conservative 83; Mismatches 216; Indels 263; Gaps 40;

QY 58 SLSPROLGFPCEAVSGSTERVELAY---ALAKNVKLTSTEQRLAHRLSEPPEDL 113  
DB 2 SISP---LKPCCL---LQSRGKKIEANRQKALARAEXKLALBQHQRPAQSKGPSNL 54  
QY 114 DALPLDL---LRF-----LNPDAF-----SGPOACTRF 138  
DB 55 PRDSKSGSHGIFKQONPSSSHGDDQRQNPHSFPPTSTEQAKGMQRPPEMPTACPSY 114  
QY 139 F--SRITKANVDL---LPRGAPERORLLPRAALACWGRGSLLSRADVRALGGLACDL-P 191  
DB 115 RPPMOVYAGISLPLANSPPGVPSQO-----LWG-----CELGO 148  
QY 192 GRFVAESAENVLPRLVSCFPLDDQOEAARALQGGPPYGP-----PST---WS 239  
DB 149 GHQASLETOSTPRANTHTHERLKVKNQETFAASSSQPRDPLEARIIGRPSTGONIS 208  
QY 240 VSTMDALRGLLP-----VLGQPIRSTIPQG-----IVANRQRSSR--D 276  
DB 209 GSVMPRTGRLQQAQAGTPMHRVGSQOGRICRNGERFQVKIGYNEALIVAFKSLPSRSYD 268  
QY 277 P---SW-----ROBERTILRRFRRE--VEKTACPSCKKAREIDESLIFY 316  
DB 269 PATTWNSMTDYPLPKAAORLPGLT-LQPLEAGEHMESSPSTSSG-----TI 316  
QY 317 KKWELACVDALLATQMDRVNAIPFTEQDLVKHKLDELVPGYRPESVIOHLGYFLK 376  
DB 317 AKTGPAAPSLAFYKGGQCVLSRARF---EADI-----SYSDLLA---LFKQ 358  
QY 377 MSPEDIRKMNVT-----LETLKALLEVDKGHEKSPQAPRRRPLPQVATLIDRFVKGGRQ 430

DB 359 M---DSRKVDYKTRKMSFLLEEYSKLMERVNG---PPQVQDLPKLTLLF-----RAQ 406  
QY 431 LDKDTLDTLTAFPEYGLCSLSPDEELSSVPPSSIMAVRQDDLDTCDDPROLDVLYP-----484  
DB 407 LQK-----TSLSP--VADLPEA-----DISRVDKSLVSSLLPFORAGV 442  
QY 485 -----KARLAFQUNNG-----SEYFVKIOSFLGAPATED-----LKAL 517  
DB 443 NFATAQRGRLLADMDGIGKTIQALICIAIYRKEMPLLVVPPSSYRFTWEQAFCHMLPSL 502  
QY 518 SQOAVSMDLAFEMKLRTDAVPLT---VAEVOKLLGPHVEGLAEERHRRPVRDMLRQR 573  
DB 503 NPLDIN-VVTKGRDLTGGLVINVSFDLSKLEQKRPFRVYIIDESE-----550  
QY 574 QDDDLTGLGLGGIPNGYLVLDLSVOE-TISGIPCLLGPVULT-VLAL 621  
DB 551 --FLKNITAVCAAMP---LLKVARVILLSGTPAMSPAEVLTQIILAV 594

## RESULT 9

US-09-722-708-55

Sequence 55, Application US/09722708

GENERAL INFORMATION:

APPLICANT: Hockensmith, Joel W.

Muthuswami, Robin

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR

TARGETING DNA METABOLIC PROCESSES USING

AMINOGLYCOSIDE DERIVATIVES

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE &amp; EDMONDS LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/722,708

FILING DATE: 28-Nov-2000

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/179,558

FILING DATE: &lt;Unknown&gt;

APPLICATION NUMBER: U.S. 60/063,898

FILING DATE: 31-OCT-1997

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 9426-005-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)7909090

TELEFAX: (212)8699741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 941 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 55:

US-09-722-708-55

Query Match 3.1%; Score 99.5; DB 5; Length 941;  
Best Local Similarity 20.8%; Pred. No. 2;  
Matches 148; Conservative 83; Mismatches 216; Indels 263; Gaps 40;



```

Db 14 LPLVLLLLAPARQAAARCPACICDNRHRYACRYONLTPVDAIPELTQRL----- 67
QY 167 CWGVRGSLSSADYRALGGLA-----CDLGRFPAESA-----EVLPRLVSCGP 212
Db 68 --DLGNLTKVTPAAAFQGVHLLHDLRCEV--ELVABGAFGLGRLLNLNLAS--NH 121
QY 213 LDDQOEA-----ANALQGGPYPGPSTW-----SVSTMDALRGILPVLAGPIIRSI 261
Db 122 LRELPGELDGLSLRLELEGNALFEELRPGTFCALGALATLNLHNLVYL--PAMAF- 178
QY 262 POGIV-AAAROSSRSDPSWROPERTILPRFR-----EVEKACPSGKAR-----EI 309
Db 179 -QGLLRVRLSLSHNALSVLAPALAGLPALRSLHNNELQALPGVLSQARGLAREL 237
QY 310 DESLIFYKME-----LEACVDALLATQMDRVNAIPFTYEQDLVKHKLDELVP-Q 360
Db 238 GHNLPTTAGEDGLALPGRLLELDGALQALGPRAFHPCRLHTLDRGNQDLTPPLQ 297
QY 361 GYPSVIOHLG-----YFLKMSPEDIRKMNVTSETLKAAL--LEVDKGHE 404
Db 298 GPGQLHQHGCIARELAGALAGLRLIYILSDNQLAGLSAAALEGAPRLGLVLERNR 357
QY 405 MS-POAPRRPLQVATL-----IDREYKG 427
Db 358 LQVPGALRALPSLSLHLDQNAVDRAPG 387

```

# RESULT 12

```

US-09-976-594-449
; Sequence 449, Application US/09976594
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT FILING DATE: 2001-10-12
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR APPLICATION NUMBER: 60/240,409
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 449
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 522433CD1
US-09-976-594-449

```

Query Match 3.0%; Score 96.5; DB 5; Length 308;

Best Local Similarity 22.9%; Pred. No. 0.69; Mismatches 126; Indels 137; Gaps 19;

```

QY 19 LPLFLSLGWHPARTLAGETGESAPLGVLTTPHNISLSPPROLGFPCEVSGLSTE 78
Db 14 MLVLVLVSWL-----PHGAL-----SLAEASRASFP--GPELHTE 49
QY 79 --RVRELAVALAOKNVKISTEOLRCLAHRLSEPPEDDALPLDLILFNPDAFSGPQACT 136
Db 50 DSRRELKRRYEDLLTRLRANQ-----SMEDSNTDLVPAPAVRILTPVRLGSGG-- 99
QY 137 RFRSRTKANDLPRGAPERO-----RLPPALACWGRGSLSSADYRALGGLACD 189
Db 100 HLHLRISRA--ALPEGLPEASRLHRLFLSPASRSMDVTRPLRLQSL----- 147
QY 190 LPRFVAESAIVLPRLVSCGPLDDO-----QEARA-----ALOG 227
Db 148 -----ARQAPALHLRL--SEPPSGDQLLAESSARPOLLEHLRPOAARRRARANG 200
QY 228 GGPYPGPSTWSVTMDALRGILPVLAGPIIRSIPOGIVAAARORSSRDPSWROPERTIL 287

```

```

Db 201 DHCLPGRCRCRLHTVRA-----SLEDLGMAD---WVL 230
QY 288 RPRRREVEKT-----ACSGKAKAREIDESLIFYKKWELEACVDALLATQMDRVNA---I 340
Db 231 SP---REVQVMTGICAGCSQFRAA-----NMHAQIKTSLRHLKRPDTPVAPCCV 275
QY 341 PFTYEOLDVLKHKLD 355
Db 276 PASYNPM-VLIQKTD 289

```

# RESULT 13

```

US-09-981-353-188
; Sequence 188, Application US/09981353
; GENERAL INFORMATION:
; APPLICANT: Laasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 188
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 522433CD1
US-09-981-353-188

```

Query Match 3.0%; Score 96.5; DB 5; Length 308;

Best Local Similarity 22.9%; Pred. No. 0.69; Mismatches 126; Indels 137; Gaps 19;

```

QY 19 LPLFLSLGWHPARTLAGETGESAPLGVLTTPHNISLSPPROLGFPCEVSGLSTE 78
Db 14 MLVLVLVSWL-----PHGAL-----SLAEASRASFP--GPELHTE 49
QY 79 --RVRELAVALAOKNVKISTEOLRCLAHRLSEPPEDDALPLDLILFNPDAFSGPQACT 136
Db 50 DSRRELKRRYEDLLTRLRANQ-----SMEDSNTDLVPAPAVRILTPVRLGSGG-- 99
QY 137 RFRSRTKANDLPRGAPERO-----RLPPALACWGRGSLSSADYRALGGLACD 189
Db 100 HLHLRISRA--ALPEGLPEASRLHRLFLSPASRSMDVTRPLRLQSL----- 147
QY 190 LPRFVAESAIVLPRLVSCGPLDDO-----QEARA-----ALOG 227
Db 148 -----ARQAPALHLRL--SEPPSGDQLLAESSARPOLLEHLRPOAARRRARANG 200
QY 228 GGPYPGPSTWSVTMDALRGILPVLAGPIIRSIPOGIVAAARORSSRDPSWROPERTIL 287
Db 201 DHCLPGRCRCRLHTVRA-----SLEDLGMAD---WVL 230
QY 288 RPRRREVEKT-----ACSGKAKAREIDESLIFYKKWELEACVDALLATQMDRVNA---I 340
Db 231 SP---REVQVMTGICAGCSQFRAA-----NMHAQIKTSLRHLKRPDTPVAPCCV 275
QY 341 PFTYEOLDVLKHKLD 355
Db 276 PASYNPM-VLIQKTD 289

```

# RESULT 14

```

US-09-897-516-6629
; Sequence 6629, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Cordin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.

```

APPLICANT: Huesing, Joseph E.  
APPLICANT: Krasomil-Osterfeld, Karina C.  
APPLICANT: Malvar, Thomas M.  
APPLICANT: Slater, Steven C.  
APPLICANT: Spiridonov, Sergei  
TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof  
FILE REFERENCE: 38-21(51847)B  
CURRENT APPLICATION NUMBER: US/09/897,516  
CURRENT FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/215, 161  
PRIOR FILING DATE: 2000-06-30  
NUMBER OF SEQ ID NOS: 8409  
SEQ ID NO 6629  
LENGTH: 689  
TYPE: PRF  
ORGANISM: Xenorhabdus sp.  
US-09-897-516-6629

Query Match 2.9%; Score 95.5; DB 5; Length 689;  
Best Local Similarity 23.4%; Pred. No. 2.8;  
Matches 91; Conservative 47; Mismatches 130; Indels 121; Gaps 21;

QY 44 APLGG---VITPHNISLSRPLGFCACVSGL---STERYRELAVALAOKNKL 95  
DB 88 AGLGGLEFALACHARISLSDKTRGLPEVOLGLPGSGGTORLRLIGVSSALDMILT 147  
QY 96 TEOLRC-LAHLSEPPEDLALPLDLFL-----NPDAFS-----GPQA 134  
DB 148 GRLKAKAKAKLGYVD---DAVPLDILDVAVOYKKVAKRKPLAWSORLLASSFGRL 204  
QY 135 CTREFSRITKAN-----VDLPRGAPERORLPLAALACWVGSLSEADVAL 183  
DB 205 LFRWQKTRAKTHGHYPAPERIIDIVIRLLEKGGK---AGL-----QAEKAF 250  
QY 184 GGLACD---LPGFVA-----ESAIVLPRLVSCGRLDQDOQEARALOGGPPY 232  
DB 251 GELAWSSSSSLRLFFATYTSKNETGSSELPAQIRHVIGLG-----GGLMGCG--- 299  
QY 223 GPPSTWSTMDALGGLPVLGQPI-IRSIPOGIYAAWRQSSR-----DPSWQPERT 285  
DB 300 -----IANTATRGHLPAKIKDINKEGINQALKYTDMISKRSQRLPSESRQMS 352  
QY 286 ILPRFRREVKTACPSGKKAREI---DESLIFKKW---ELEACVDALLATQMDRVN 338  
DB 353 IL-----SGTIDYSGEQADIVAEVFEEDLVLRKMAVEAEANTKAETIFA--SMTS 402  
QY 339 AIPFTYEQDLVKKHLDLIPQGYPEYI 367  
DB 403 SLPI-----HOIAETIAKR--PEOVI 420

RESULT 15  
US-09-815-242-5107  
Sequence 5107, Application US/09815242  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5107  
LENGTH: 4342  
TYPE: PRF  
ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5107

Query Match 2.9%; Score 95.5; DB 5; Length 4342;  
Best Local Similarity 22.2%; Pred. No. 42;  
Matches 89; Conservative 61; Mismatches 149; Indels 101; Gaps 21;

QY 214 DODQEARALOGGPPYGPSTWSTMDALRGLFVLCOPITRSTPOGIVAAWRORS 273  
DB 1579 EAEQORLRAALQAEPLPYWPT---OLMRLAQWPLPGSKLDTRALPEPV---WQORE 1631  
QY 274 SRDPS-----WRQP---ERTILPRF-----RREVEKTACSGKKAREID 310  
DB 1632 HVEPTELORRIAAIWSEVGLPRVGLKDDPELGSHSLATRIYSRT---RACDVE 1686  
QY 311 ESL-IFYKKWELEA-C--VDALALA-----TOMDRVAIPFTYEQDLVKKHKLDEL 357  
DB 1687 LPLRALFEASELEACEQVRAQAAGRTDSHGAIRRIDREQPVPLSYSQ---QRMWFL 1741  
QY 358 YPQGYPEVIOHLGLYFLKMSPEDIRKNNVTSLETLKALLEVDKCHENSPOAPRRPLQV 417  
DB 1742 W-OLEPDSPAYNVGGLARLSGLDVARFEA---ALQALVQ---RHE---TLRTTSPSV 1789  
QY 418 ATLIDRFYKGGOLDKDLDTLTAFTYPGYLCSLSPSELSSVPSSIMAVRPDDTCDPR 477  
DB 1790 DGVPRQVRHGDGLHMDQD-----FSALDRDSROOHLQTLADSE--AHRFFDLES--- 1838  
QY 478 QLDVLYPRARLAFOMMNGSEYFVKIOSFLGAPTEDLKALSQOQVNSMDL-----AT 528  
DB 1839 -----GPLLRVOMVMKABREHLYV-----TLHHIYTEGWAMDIFARELGALYEA 1883  
QY 529 FMKLRITDAVLPIYAEVOKLGLGPHVEGILKAERHRPVRDW 568  
DB 1884 FLDRSEPLEPLPYQYLDYSWQR-EMLESGEROROLDYW 1922

Search completed: December 7, 2001, 09:49:36  
Job time: 94 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:52:27 : Search time 33.02 Seconds  
(without alignments)  
1448.747 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALQRDLPCWCGSDRPSLL.....LLGPPVLTVALLASTLA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR\_68:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1708	52.4	625	2 JC7362	Erc protein - rat
2	125.5	3.8	2109	2 T31352	hypothetical prote
3	120	3.7	2218	2 B84683	hypothetical prote
4	111.5	3.4	4957	2 T03455	ALR protein - huma
5	111.5	3.4	5262	2 T03454	hypothetical prote
6	110.5	3.4	340	2 B96632	salivarycn Mx1 syn
7	110	3.4	1770	2 T18551	hypothetical prote
8	109	3.3	1171	2 T35548	hypothetical prote
9	108.5	3.3	652	2 G82962	hypothetical prote
10	108.5	3.3	1397	2 T51292	Dna2p - fission ye
11	108.5	3.3	1398	2 T39568	hypothetical helic
12	108.5	3.3	2361	2 T25752	hypothetical prote
13	108	3.3	283	2 E82662	ABC transporter AT
14	107.5	3.3	1288	2 T42756	5-oxoprolinase (AT
15	107.5	3.3	3573	2 S23070	erythronolide synt
16	106.5	3.3	653	2 A46362	amyloid precursor
17	106.5	3.3	1748	1 JN0786	integrin beta-4 ch
18	106	3.3	391	2 T36311	probable lipase -
19	106	3.3	527	2 T75503	probable fibronect
20	106	3.3	690	2 S35251	probable membrane
21	106	3.3	1121	2 A82809	exodeoxyribonuclea
22	105.5	3.2	1607	2 T13250	hypothetical prote
23	105	3.2	617	2 F75484	serine/threonine-s
24	105	3.2	693	1 A41090	conserved hypobeth
25	105	3.2	830	2 E83031	ankyrin-like prote
26	105	3.2	1058	2 D82654	probable metallopr
27	105	3.2	1096	2 T08619	hypothetical prote
28	105	3.2	2298	2 T49648	hypothetical prote
29	104.5	3.2	264	2 S75053	hypothetical prote

30	104	3.2	503	2 A83027	hypothetical prote
31	104	3.2	1003	2 B71469	glycine--tRNA ligase
32	104	3.2	7576	2 T17428	FK506 polyketide s
33	103.5	3.2	637	2 A75342	hypothetical prote
34	103.5	3.2	1402	2 S75938	chemotaxis protein
35	103.5	3.2	4128	2 JC6306	protein kinase (EC
36	102.5	3.1	436	2 T28066	hypothetical prote
37	102.5	3.1	733	2 H69411	cell division cont
38	102.5	3.1	1010	2 T36383	probable large ATP
39	102.5	3.1	1239	2 G71266	probable ATP-depen
40	102	3.1	1159	1 H64089	DNA-directed DNA p
41	102	3.1	1160	2 S34809	avrBs3-2 protein -
42	101.5	3.1	836	2 T30312	plin biosynthetic
43	101.5	3.1	2472	2 E83594	still frame shift p
44	101	3.1	677	2 G02540	nucleobindin - hum
45	101	3.1	1997	2 T30874	virg1n1amydin S sy

## ALIGNMENTS

RESULT 1  
JC7362  
Erc protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 31-Dec-2000  
C:Accession: JC7362  
R:Yamashita, Y.; Yokoyama, M.; Kobayashi, E.; Takai, S.; Hino, O.  
Biochem. Biophys. Res. Commun. 275, 134-140, 2000  
A:Title: Mapping and determination of the cdna sequence of the Erc gene preferentiall  
A:Reference number: JC7362  
A:Accession: JC7362  
A:Molecule type: mRNA  
A:Residues: 1-625 <YAM>  
A:Cross-references: DDBJ:D87351  
C:Comment: This protein, as a tumor marker for renal cell carcinoma, is involved in c  
C:Genetics:  
A:Gene: Erc  
A:Map position: 10q12

Query Match	52.4%	Score 1708;	DB 2;	Length 625;
Best Local Similarity.	55.7%;	Pred. No. 3.1e-114;		
Matches 351;	Conservative 81;	Mismatches 184;	Indels 14;	Gaps 5;
QY	1	MALQRDLPCWCGSDR--PGSLLFLFLSLGWNHPRKTLAGEGTGSAPLGGVLTTPPHNIS	57	
DB	1	MALPRAQPLDSCGSPICSRSLFLLLSLGWLPLDQTTTTSQEAALLHAVTGTV-DFA	59	
QY	58	SLSPRQLGFCAEVSGSLSTERRELAVALAOKNKLSTBOELAHRLSE--PREDDA	115	
DB	60	SLPTLFLGLTCDDEVSGLSMGAKELAMAVKOKNKLVLQYHQLRCARLRPKLTLTEEDDA	119	
QY	116	LPDLFLFLNPDAFSGPQACTFEFSRITKANVDLPRGAPERQRLPLPALACWGVRSLL	175	
DB	120	LPDLFLFLNPDAFSGPQACTFEFSRITKANVDLPRGAPERQRLPLPALACWGVRSLL	179	
QY	176	SEADVRALAGLACDLPGRFVAESAELLPLRVSCGCPDLDDQQAARALDGGGPPYGP	235	
DB	180	SETDARALGGLACDLPGRFVAESAELLPLRVSCGCPDLDDQQAARALDGGGPPYGP	239	
QY	236	STWASTMDALRGLLPVIGOPILIRSIPOGIVAMRQRSSRDPSPQPERTILRPFRRREV	295	
DB	240	STWASTMDALRGLLPVIGOPILIRSIPOGIVAMRQRSSRDPSPQPERTILRPFRRREV	299	
QY	296	EKTACPSGKKAKEIDSLFFKKWLEACVDALLATQMDRVNNAIPFTYEQDLVLKHLKD	355	
DB	300	EOKACPPGKEPVDNDENLFFYQNMELACVDOTLLAGOMDLVNEIPFTYEQSLIRKHKLD	359	
QY	356	ELIPGPGYPSVYQHGLYFLFKMSPDIDKKNVTSLETETKALLEVKGKHEMSQAPRRRLP	415	
DB	360	KTYPGYPSVSLIKQGHFFRYVSPEDIRQWNTSPDVTNLTLSKSGKMD-----A	411	

```

Oy 416 QVATLIDREFVKGQGLDKDITLTAFYPGYLCSLSPBELSSVPPSSIWAVRPDLDTCD 475
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 412 QYIALVACVLRGGKGLDEIVAKLDINIPILSLYLCDFSPQHLAIPSSVMVLVGLHDLKDS 471
Oy 476 PQQLDVLTPKALATONMNGSEFYFKIOSFLGCAPEDELDKALSOONVSDLTATPKLRPD 535
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 472 QHHLGILYOKACSAFONVSGLEFYFKIRTEFLGASREDIRALSOHNVSMDITFKKMOVD 531
Oy 536 AVLPTVAEVOKLGPVHEGLKAERHRPVRDMILRQRODDLTGLGLOGGIPNGYLV 595
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 532 ALVGSVAEVOGLLPHIGDLKTEDEKSPVRMDLFRQOQKDLDSLGLGLOGGIPNGYLV 591
Oy 596 DLISVETLSGTCCLLGPVLTVALLLAS 625
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 592 DFNVRFAFSSGAPLLGPGFVFAMIPALLSA 621

RESULT 2
h31352
hypothetical protein - Pelargonium x hortorum
C:Species: Pelargonium x hortorum
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: t31352
R:Downie, S.R.; Katz-Downie, D.S.; Wolfe, K.H.; Calie, P.J.; Palmer, J.D.
Curr. Genet. 25, 367-378, 1994
A:Title: Structure and evolution of the largest chloroplast gene (ORF2280): internal pla
A:Reference number: 221012; MUID:94363755
A:Accession: t31352
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2109 <DOM>
A:Cross-references: EMBL:M83200; NID:g468913; PID:g468914; PIDN:AAA73173.1

Query Match 3.8%; Score 125.5; DB 2; Length 2109;
Best Local Similarity 20.5%; Pred. No.2.2; Mismatches 207; Indels 199; Gaps 32;
Matches 128; Conservative 91;

Oy 79 RVELALVALAOKNVKLSLEQRL-----CLAHRLSEPPEDLALPLDLLFLNPDASFQ 133
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1006 RIREIS-EMCLRNLTLSAERIRRNESPLTHLRSPNVLEFLYSTLLLL-----VAGTL 1060
Oy 134 ACTREFSRITANDL-----LPGKAPERORLL-----PALACWGVKSLSEA 178
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1061 VCT-YLSRLSKDYGLQELKVKVSLMIPSYTIELRLMDRYPSELNFGIKNFIYAM 1119
Oy 179 DVRALGLACLDPRGFVAESAELLPLRVSCGPRLDQDOEAAKALOGGPPYPP----- 235
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1120 EE-----LKESSLVYGNM-----LEGGRATVEVAIF 1146
Oy 236 STWSVSTMDALRGLLPVLGQPIRSIPQIVAAWRORSSRDPSPWRQPERTILRPFRREV 295
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1147 SNMNLNID-ISDLISLIPNIDR-ITFSINTRLHSHTSKE-----IYSFIRKRR----- 1195
Oy 296 EKTACPSGKKAREIDSLIFKKW-----ELEACVDAALLATOMDRVNAIPF-----TYEQ 346
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1196 -----VYGAMIDKIESLSTSVADIDCDRGNLQFSTLLTTEKG 1236
Oy 347 LDVLKHKDELTPGYSPESVIOHGLYFLKMSPEDIR-----NNTSLETALKALE- 398
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1237 VDQILSLTOSKSNAGSGOMTEQPEMTLR-HVVDLQKKYLMGEFNSSLAERRIFLAH 1295
Oy 399 -----VDKGHMSPOAP---RRPL---POVATLIDRFVYGRQLDKD----- 434
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1296 YQTMYSKTCGCVNAFHRPSHEKPFSPRLDSPRGILVIGSIGRGYLLKSLAKNHF 1355
Oy 435 ---TL-----DTLTAFYPGYLCSLSPBELSSVPPSSIWAVRPQOLD 472
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1356 PLITLEMEARSSWPFQHLIDEIYGDQLEYVDYDTSLSVEVEE-----EEDTSMGIEWELP 1411
Oy 473 TCDDROLDVLTPKARL-AFOMMNGSEYFVKIOSFLGCAPEDELDKALSOONVSDLTATPK 531
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1412 -DPRDEDEIDQAEKMDTRRLDGIETTHAIQDMT-----DL-----GISVDAQ--LN 1455

```

```

Oy 532 LRTDAVLPTVAEVOKLLGPVHEGLKAERHRPVRDMILRQRODDLTGLGLOGG----- 587
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1456 FLPESTL-----INEIEGHVDLDEDAETEL-----WELMEEMDRHLGLSLQFALYRA 1505
Oy 588 -----IPNGYLVLDLSVOETLSG 605
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1506 MTPCIIWIPINHVD-LLEDRTTLAG 1529

RESULT 3
B84683
hypothetical protein At2g28300 (imported) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84683
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Koffac, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84683
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2218 <STO>
A:Cross-references: GB:A8002093; NID:g4803953; PIDN:AAD29825.1; GSPDB:GM00139
A:Gene: At2g28300
A:Map position: 2

Query Match 3.7%; Score 120; DB 2; Length 2218;
Best Local Similarity 19.1%; Pred. No.5.9; Mismatches 222; Indels 270; Gaps 32;
Matches 135; Conservative 81;

Oy 30 HPARTLA-----GETGESAPLGLVLTTPHNISLSPROLLGPPCAEVSGLSTERVRE 82
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 169 HPTSLALTPSPDLSGPPGFSPL-----ASPAPTPIRGRGRSGRSGARRR 217
Oy 83 LAVALAOKNVKLS--TEQRLCLAHRLSEPPEDLALPLDLLFLNPDASFQACTRFFS 140
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 218 VEGVHGSNSSTTQRTERTATSLA-----SDAEA-----TRFALPRASSETVS 259
Oy 141 RTKANVLLPRGAPER-ORLLPALACWGVKSLSEADYALGLACDLPGRVAESA 199
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 260 RVPKMN-----EGTSNPDQVSPVHSATTLALRSKAAKDL-----DAPGDSGSH 306
Oy 200 EYLLPRLVSCPRLDQDOEAA-----RAALOGGPPYGPSPSTWSV-----STMD 244
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 307 -----VQTLNVLENSSEKRAFAVKRRLIQQGGPQNONAVSSVCGSKSPSEGRYYT 359
Oy 245 ALRGL-----LPVLGQPI-----IRSIPOGIYAAMQR----- 272
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 360 ALQGVTTAPSDATLPMSSQPSDATLPMSSQPSGVTEAQAENVPSLPALPAKRRVRNLP 419
Oy 273 SSRDPSWPPER-----TIL-----RPRRR 293
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 420 SRGETPKRQGRROPPLPATDASSARSTGLIPOLEVKKVNGSITKAKADAVAKQEPHESQ 479
Oy 294 EYEKTACPSGKKAREIDSLIFKKMELEACVDAALLATOMDRVNAIPFTEQOLDVLKHK 353
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 480 SVAPDIHSSGSLQGIIRDGTSGTGSARKQYADYDVARVKKEI-----PSESLKHK 533
Oy 354 LDELTPQGYSPESVIOHGLYFLKMSPEDIRKMNVTSLFTLK-----ALLEVDKG 402
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 534 VGE--PSATYTRTNVPD-----AOSPGEH---NLHTVETHKAEDSSGLKNOEALYNLSKA 582
Oy 403 HEMSPQARPRLPQVATL-----IDRF 424
    |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 583 DKLSVDIP-HVYFGDLTSSGVANKDVIGSSKVAENELVYKIRGGVDSSVYQLSLANT 641
Oy 425 VKGRQOLDKDTLDTL-----TAFYPGYLCSLSPBELSSV-----PPSSIWAVR 467

```

```

Db      642 LTAASLEKCTADQLGEBKLSQEGETTPASDGETCHLAEATSSLSYVRSEPTASA----- 697
Oy      468 PODDTCDDPRLDVLVYRKARLAFON-----MNGSEYFKIQSLGALPTEDLKALISOONVS 523
Db      698 ---STAEPLPTDRL--EKNISFODEVKTLNGD-----KREAIL-----LSSEQOTVWN 741
Oy      524 MDLAT-----FMKLRDVAVLPLTVAEVOKLGLGPHVEGLKAEERHNPVRD 567
Db      742 SKITENSEELQASTDEY-----PHYDGKSVDAVNAQTVKE 776

```

## RESULT 4

```

T03455
ALR protein - human
C:Species: Homo sapiens (man)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03455
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A:Reference number: Z14954; MUID:97388474
A:Accession: T03455
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4957 <PRA>
A:Cross-references: EMBL:AF010404; NID:g2358286; PID:AC51735.1; PID:g2358287
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

```

```

Query Match      3.48; Score 111.5; DB 2; Length 4957;
Best Local Similarity 19.9%; Pred. No. 74;
Matches 148; Conservative 112; Mismatches 287; Indels 195; Gaps 34;

```

```

Oy      8 PCWCGDRPGSLFLFLSLGWVNP--ARTLAGETGESAPLGVLTPPHNISLSRQLL 65
Db      3557 PSVSLGDPGSMTONL--LGFOQPMLEKPMONNTGPPRKGPVLQSGGLPGVGINPTV 3614
Oy      66 GFPCAENVSG-----LSTERYRELAVALAOKNVKLTSEOLKSLAHRLSEPPD--L 113
Db      3615 GOLRAQOLGVLAKNPQLNHLSPQOOQOLQALIMORLOQSO-----AVROTTPYQEBGT 3668
Oy      114 DALPLDLLFLNRP--AFSGQACTRFSSRTITKANVLLPRGAREKQYLLAALACWYR 171
Db      3669 QTSPLQGLGQPOLGGRPGPOT---GPIQELGAGRPQG--PRILRAPGALSTGFLV 3722
Oy      172 GSL-----LSEADVRAALGGLCDLPGRFVAESAELL-----PRIVSC 209
Db      3723 GVNPTRPPSSPQEBKR-----PSQLPSPSSQLPTREAOPLRTNHGTRPKQSGTLEPP 3774
Oy      210 PGLPDQDOEARAALAGGGRPYGPRSTWSVTMDALRGLLP-----VLGQPIITRSIPQG 264
Db      3775 PGRVSPAQAQADLTFLSKGLRPMDBRPDLAETQKPEQSSLVPGHLDYVNGO-----VVPEA 3830
Oy      265 IVAAMRQSSSDPSMRQERTILRRFRREVE-----KTACSPGSKKAREIDE 311
Db      3831 SOLSTKQPREPC-----ALGAOSVREANGERTIGARGTSMHLLLAGPSEAGHLLQ 3884
Oy      312 SLIFPKKMLACVDAALLATQMD--RVNAIPFTYEOLDVLLKHKIDELTPQGYPREVIOHL 370
Db      3885 KILRAKNVQLSTGGSEBLRAEINCHIDS-----KLAGLQKL-----QGTSPSKEDAA 3933
Oy      371 GYLFLKMSPEDIRKWN---VTSLETLK-----ALLEVDKGHEMSPPQARRPLPOVA 418
Db      3934 ARKPLTPKPKHVQKASDRLVSSRKLRKEDGVRASEALLK-----QLKQELSLPLTPEPA 3988
Oy      419 ---TLIDRIVVG-----RGQDKDLDLTITATFYPTVL---CSLSPEELSSVPP 460
Db      3989 ITANFSLPAPGSGCPVNGQSOLRGAFSGGALPTGPDYYSQULLTKNNLSNPPTPPSSLP 4048

```

```

Oy      461 SSIMAVRPODDTCDDPRLDVLVYRKARLAFON-----MNGSEYFKIQSLGALPGALP----- 510
Db      4049 TPPSVQOKMWNVCTPSEELGHNKDAASADSERALROTSE--VKSILDLAALPTPHN 4106
Oy      511 -TEDLKALISOONVMDLATEFMKLRDVAVLPLTVAEVOKLGLG-----PHYGLKAEERH 563
Db      4107 QTEDEVREMSDESD-----SPDSIVPASPE--SIIGEAPRFPHLGSGRWGEDR 4155
Oy      564 FVRDMILRQDDLL---DT-----LGIGLOGGIP-----NGVYLVDLSOET 602
Db      4156 ALSPVILPIRDSIPVPEDTPYGALGLEVPGKLPYTWKEKGKSEVSVMLTVSAADKN 4215
Oy      603 LSGTRCLLGPVPLVALLLA 624
Db      4216 LNG-----VMVAVALLS 4228

```

## RESULT 5

```

T03454
ALR protein - human
C:Species: Homo sapiens (man)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03454
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Ya
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homo
A:Reference number: Z14954; MUID:97388474
A:Accession: T03454
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5262 <PRA>
A:Cross-references: EMBL:AF010403; NID:g2358284; PID:AC51734.1; PID:g2358285
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

```

```

Query Match      3.48; Score 111.5; DB 2; Length 5262;
Best Local Similarity 19.9%; Pred. No. 81;
Matches 148; Conservative 112; Mismatches 287; Indels 195; Gaps 34;

```

```

Oy      8 PCWCGDRPGSLFLFLSLGWVNP--ARTLAGETGESAPLGVLTPPHNISLSRQLL 65
Db      3662 PSVSLGDPGSMTONL--LGFOQPMLEKPMONNTGPPRKGPVLQSGGLPGVGINPTV 3919
Oy      66 GFPCAENVSG-----LSTERYRELAVALAOKNVKLTSEOLKSLAHRLSEPPD--L 113
Db      3920 GOLRAQOLGVLAKNPQLNHLSPQOOQOLQALIMORLOQSO-----AVROTTPYQEBGT 3973
Oy      114 DALPLDLLFLNRP--AFSGQACTRFSSRTITKANVLLPRGAREKQYLLAALACWYR 171
Db      3974 QTSPLQGLGQPOLGGRPGPOT---GPIQELGAGRPQG--PRILRAPGALSTGFLV 4027
Oy      172 GSL-----LSEADVRAALGGLCDLPGRFVAESAELL-----PRIVSC 209
Db      4028 GVNPTRPPSSPQEBKR-----PSQLPSPSSQLPTREAOPLRTNHGTRPKQSGTLEPP 4079
Oy      210 PGLPDQDOEARAALAGGGRPYGPRSTWSVTMDALRGLLP-----VLGQPIITRSIPQG 264
Db      4080 PGRVSPAQAQADLTFLSKGLRPMDBRPDLAETQKPEQSSLVPGHLDYVNGO-----VVPEA 4135
Oy      265 IVAAMRQSSSDPSMRQERTILRRFRREVE-----KTACSPGSKKAREIDE 311
Db      4136 SOLSTKQPREPC-----ALGAOSVREANGERTIGARGTSMHLLLAGPSEAGHLLQ 4189
Oy      312 SLIFPKKMLACVDAALLATQMD--RVNAIPFTYEOLDVLLKHKIDELTPQGYPREVIOHL 370
Db      4190 KILRAKNVQLSTGGSEBLRAEINCHIDS-----KLAGLQKL-----QGTSPSKEDAA 4238
Oy      371 GYLFLKMSPEDIRKWN---VTSLETLK-----ALLEVDKGHEMSPPQARRPLPOVA 418

```

```

Db 4239 ARKPLTPKPKRVOKASDBLVSSRRKLRKEDGVASEALLK-----OLKQELSLPLTEPA 4293
OY 419 -----TLIDRFYKKG-----RGQLDKDITDITLTAFFPGTL-----GSLSEELSSVPP 460
Db 4234 ITANFSLAPFESGCPVNGQSOLRGAFSGALPTPDYYSQLLTNNLSNPTPPSSLP 4353
OY 461 SSIMAVRPQDDTCDPROLDVLYPARLAFON-----MNGSEYFVKIOSFLGAP-----510
Db 4354 TPPEVQOKMNGVTPFSEELGHPKDAASARSEKALDTSF--VKSIDLALALTPPHN 4411
OY 511 -TEDLKALSOQNVSMDLATFMKLRDVAVLPTVAEVQKLG-----PHEGLKAEERRH 563
Db 4412 QTEDEVMSDESDSD-----SPDSITVPASSPE--SILGEADRFPHLGSGRWEQEDR 4460
OY 564 PYRDMILNQRODDL-----DT-----LGGLGLOGIP-----NGYLVLDLSVOET 602
Db 4461 ALSVPILPIPRDSIVFPDTPKRYGALGLEVPGLKLPVTWEKGSSEVSMVLSAADKN 4520
OY 603 LSGFPCLLGPGLVTLVALLLA 624
Db 4521 LNG-----VMYVAVALLS 4533

```

## RESULT 6

```

hypothetical protein F8A5.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: B96632
R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huitzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malli, R.; Matzidali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B96632
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-340 <STO>
A:Cross-references: GB:AE005173; NID:g2462761; PIDN:AB71980.1; GSPDB:GM00141
C:Genetics:
A:Gene: F8A5.20
A:Map position: 1
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

```

Query Match 3.4%; Score 110.5; DB 2; Length 340;
Best Local Similarity 20.4%; Pred. No. 2;
Matches 89; Conservative 62; Mismatches 126; Indels 159; Gaps 22;
OY 87 LAQKNVLTSTQRLCIA--HRLSEPPEDLDLPLDLLFLNPDASFQOACTRFSRITK 144
Db 3 LGSQGLEVSAGGLGCMALSAHYGAKPPTDAIAL-----LHNAIISG-----VTFP-----48
OY 145 ANVDLLPGABERORLLPALALACWGR-----GSLSEADYRALGLACDLPGRFV 195
Db 49 --DTSMDYGETNELLLGKALKDGVKEKVELATKFGFFIVEGISEVRG-----DPEYV 100
OY 196 AESAEVLLPRL-VSGPG-----PLDQDQEAARALOGG-----GPPYGPST--237
Db 101 RAACEASLRDLACIDLYYQHRIDTRVPIETMRELKLVKEEGIKYIGLSEASASTIR 160
OY 238 -----WSVTMDALRGILPV-----IGOPITRSIP--OGIYAANRO--RSSRD 276
Db 161 RAHAVHPTTAQVIEKSLMSRAEEDITPICRELIGIYAVSPLGFLAAGPKLAENLEN 220
OY 277 PSHQPERTILRPFRREVERTACPSGKAREIDSLIFYKK---WELACVDAALLAT 332

```

```

Db 221 DDFRATATTL--PRQOE-----NVDNKILFEKVSAMARKGCTPQALALA 265
OY 333 -----QMDRYNAIPFTEQDLVKKHKLDELTPGCEYSYQHLGTYFLKMSPEIDIKKMYT 388
Db 266 WVNHOQDDVCPITGTTK-----TENNONRALRSVKLTPEEI-----S 303
OY 369 SLEETKALLEVDKGHMSPOARPRPLQVATLIDRFVKGRLDKDITDITLTAFFPYIC 448
Db 304 ELDSL-----APES-----VAGE-----RIMA 321
OY 449 SLSPPELSSVPPSSIM 464
Db 322 SMSTFKNSNTPLSSM 337

```

## RESULT 7

```

T18551
saframycin Mx1 synthetase B - Myxococcus xanthus
C:Species: Myxococcus xanthus
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T18551
R:Pospiech, A.; Bietenhader, J.; Schupp, T.
Microbiology 142, 741-746, 1996
A:Title: Two multifunctional peptide synthetases and an O-methyltransferase are invol
A:Reference number: Z18967; MUID:97090395
A:Accession: T18551
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1770 <POS>
A:Cross-references: EMBL:U24657; NID:g1171127; PID:g1171128; PIDN:AAC44128.1
A:Experimental source: strain DM504-15
C:Genetics:
A:Gene: safB
C:Superfamily: acetate--CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein
F:72-577/Domain: acetate--CoA ligase homology <ACL>
F:618-668/Domain: acyl carrier protein homology <ACP1>
F:1222-1659/Domain: acetate--CoA ligase homology <ACL2>
F:1678-1746/Domain: acyl carrier protein homology <ACP2>

```

```

Query Match 3.4%; Score 110; DB 2; Length 1770;
Best Local Similarity 23.3%; Pred. No. 22;
Matches 111; Conservative 49; Mismatches 175; Indels 142; Gaps 22;
OY 1 MALQRLDPCWCSGDRPGSLFLFSLGWVHVR-----TLAGET---GIES 43
Db 520 VVOGEVDRRYRPGDWPDIYAIRDISFOHARHVAVLLKSGLSLTKTSSGKVGQATRE 579
OY 44 APILGVLT-----PHNISSLSPRQLGFPCE-----VSGLSTERRV--ELAVALA 88
Db 560 AYLBQDLTVSADAQGEVGEFSLSRASLIERSEKERPALVAYIYASATYAGSIPRAIL 639
OY 89 QKNVLTSTQRLCIA--HRLSEPPEDLD-ALPLDLLFLNPDASFQOACTRFSRIR 142
Db 640 TGEVRLPOLGDSLALVELKRIE---ODLEVALLETRLL-----EGPTLGAVALHL 689
OY 143 TKANVDLLPRGAP-----ERORLLPALALACWGRKSLSEADYRALGLACDLPGRFV 197
Db 690 EAAGREARPVASPPGTGERLPVPPGORALW-----FLHOLADSDPGYTA- 735
OY 198 SAEVLLPRLVSGPGLDDQEAARALOGGPGPYGPSTWSVTMDALRGILPVIGOP 257
Db 736 -----RAICLOGALD--VEALRRAFOS-----LVTRHPLRATFPVAGTEP 774
OY 258 IRSIPQGIYAAMR--ORSSRDSWROPERTILRPFRREVERTACPSGKAREIDSLIFY 316
Db 775 VOQVAHVPASFLVVDASG-----DEALARRLLRRA-----Y 808
OY 317 KKWEL-----EACVDALLATQMDRYNAIPFTEQDLVKKHKLDELTPGCEY 364
Db 809 RPFIDGRPLRTHLFSRAQEHVLLS---VHHIVDFWSLAVLVDLRLRYEAGGDG 864

```

```
Qy 365 SVIOHGLYFLKMSPEDIRKWNVTSLETLALLEVNDKH--EKSPOAPRRPLQVA 418
   | : | | | : | | | : | | |
Db 865 VLLPVG-----QPADAIRWQAEYMGPSH--EEDMGYWRQLAGPLRLLELPAPA 913
```

**RESULT 8**

hypothetical protein SC6G3.04 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35548  
R:Seeger, K.J.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL data library, July 1999  
A:Reference number: 221582  
A:Accession: T35548  
A:Status: preliminary; translated from GB/EMBL/DDBT  
A:Molecule type: DNA  
A:Residues: 1-1171 <SEE>  
A:Cross-references: EMBL:AL009682; PIDN:CABA6976.1; GSPDB:GN00070; SCOEDB:SC6G3.04  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC6G3.04

Query Match	3.38;	Score 109;	DB 2;	Length 1171;
Best Local Similarity	20.78;	Pred. No. 14;		
Matches 163; Conservative	69;	Mismatches 246;	Indels 310;	Gaps 35

OY	3	LORL-----	DPWCSGGRPSLSLFLPSLGM---	VHPARTLAGEPTESAPLPG--	GYL	50
Db	402	LORLKEGCVGAETAEVEVETGAGESSALTTRKRAAMTSPVAARLDLVGVRGVTAAQAADGTL				461
OY	51	TTPH-----	NISSLSPROLLGFP--CAEV---	SGLS--TERVRELAVALAO-----		89
Db	462	RENHRAAEGRVTSARVVALGMAARCAFLTDLHDGTLTEAEVREPGAALPELLAALDL				521
OY	90	-----	KNVKSTQQLRCLAHRL-----	SEPPEDLDA-----		115
Db	522	LDSVRRRLPETETVTRIRARLGLLLDAVRLPGLAGSDPEPDAVVVTLAVRSAAD				581
OY	116	--LPDL-----	LLFLRN-----	PDAFSGPOACTRF		138
Db	582	RLGLRLDGELYALSRGSPLLQGAQAARVLLDDOSDLGTRFLAGWVPTANGPGRRL				641
OY	139	FSRTTKANY-----		DLPRGAPERQ		158
Db	642	ERRRLTVLVAAGFLIESASTALGPREFERVETMSDRGFLDRLYLALRGFALRPQO--	RS			698
OY	159	RLRLPALACMCVRCSSLSSEADVNRALGSLDLDGRGVASAEVLLPRIVSCGPRLD--	QD			216
Db	659	RYLVAASDRCLDRDRLPRAPAEIVGIMAA--ADGEGALLGELGLADLASAPRAADACAT				757
OY	217	QOEARAALQO-----	GGPPYGPPESTVSTMDALRGLIPVLOGPIRISIP-----	OG		264
Db	758	SAPAPRTAPSCGAEPVHGAGAPRLGPRADRMRLLLGRTAG--LPALALRYARALDELFDREG				816
OY	265	IYAAWROBSND-----		PS--WQOPERTILPRPF		291
Db	817	EEADBESEKTEWEGDGEPEPGAGAGTSEGSDDDRGTGGAARSFVSVMNMEADRTLFLGAEI				876
OY	292	RREVEKTAQSPGKKARIDESLIFUKKMWELACVDUALATODRDNALPFIYEODLYLK				351
Db	877	ROEVLERRAVADGRK-----	DYALDLDPASVSPVELLSAV-----			911
OY	352	HKLDLEYPOGVESVIOHLYFLKMSPEDEIRKMNTYSLE--TKALLLEVDKGHEMSPOAP				410
Db	912	-----	LTLAGMPEQGRVASLRL--VKRLVEBELKELATRLRPLTGL-----	TPRPT		958
OY	411	RRP-----	LPV--ATL--IDRVLKKGQOL--DKOTLDTITATYRGYICLSLSPBELSVRP			459
Db	959	RRPGPDLPTLRLANLAHLRRREDGVEVEVRRPFRRTTARNDRMRLILVUDVASNE				1016

QY 460 PSSIWAVRQDDELTCDPRQLDVLYPKARLAEQNMNGSEYFKIKQSFEGAPT----- 511

```
Db      1019 TSVVMS-----ALTAAILGAPLSTHETLF 1044
```

QY 512 ----EDLKALSOQNVSM-----DLATFNKLR--DAVLPLTVAEVQKLLGP 551

Db 1045 STGVADLTGLVADPLSLLEVKVGGSTHIAAGLAHARSLVTPRDLVVVVSDFEE--GA 1102

QY 552 HVEGLKAE 559

Db 1103 AVEGLAE 1110

## RESULT 9

Hypothetical protein PA5464 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
 C1.Species: *Pseudomonas aeruginosa*  
 C1.Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C1.Accession: G622962  
 R1.Stover, C.K.; Plam, X.O.; Ertyn, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A1.Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa  
 A1.Reference number: A82950; MUID:20437337  
 A1.Accession: G82962  
 A1.Status: preliminary  
 A1.Molecule type: DNA  
 A1.Residues: 1-652 <STO>  
 A1.Cross-references: GB:AE004959; GB:AE004091; NID:99951791; PIDN:AMG08849.1; GSPDB:GN  
 A1.Experimental source: strain PA01  
 A1.Genetics:  
 A1.Gene: PA5464

Query Match	3.38	Score 108.5	DB 2	Length 652
Best Local Similarity	22.68	Pred. No. 6.8		
Matches 98	Conservative 44	Mismatches 143	Indels 149	Gaps 19

```

0Y 12 LIFSJGWHVPAHTLAGEFTEGSAPLGVLTTPHNNISSLPOLLGF----- 67
   || | | | | | | | | | | | | | | :||| :
Db 177 LIDLAGM--PORT-GONGREAHNGFVLRSPDYELSNPREFAVNNEYFLDPSTACR 232
                                           :||| :
0Y 68 -----PCAE-----VSGLSTERVELAVALAQNVK 93
   || | | | | | | | | | | | | | | :||| :
Db 233 RPALRYRYQGRGNAPBHSACAQSFAYLNAGHDGQQPLOGDPERVEVDYLLEANDN 292
                                           :||| :
0Y 94 LSTE---QLR---CLARLSEPPEDDALPLDLLFLNPDAFSGPOACTFFSRITKAN 146
   | | | | | | | | | | | | | | | | | | | | :||| :
Db 293 LVSRMGHTMLRLVICAPRRPRGPCRLD--LDQHLYLVSTAFVG----- 334
                                           :||| :
0Y 147 VDLIRPGAPERQRLLPALACM-GVRGSLSLEADVRALGSLACDPGRFVASAEVLPR 205
   | | | | | | | | | | | | | | | | | | | | :||| :
Db 335 -DL-----OLSMWDGLTGAYPSKLFVLP-----SOVIEYTKVELRS 371
   | | | | | | | | | | | | | | | | | | | | :||| :
0Y 206 LVSCGPPLDODO----OEARAALOGCGRPYGPSTWSVTMDALRGRLPVLGOPITRS 260
   | | | | | | | | | | | | | | | | | | | | :||| :
Db 372 LASVPLKTDREVASIVERAASHMSYDGYYFFISNNCAVELIKLNRGIFRRPLQSIDS 431
   | | | | | | | | | | | | | | | | | | | | :||| :
0Y 261 I-POGIWAMORSSRDSPROPERTILRPFRR-----VEKTACPGR-- 304
   | | | | | | | | | | | | | | | | | | | | :||| :
Db 432 ITPYVULEMLENRKLADSPVLDPPREALRLTGRFDSFRDYOAMEDVLKQRLHPDQVE 491
   | | | | | | | | | | | | | | | | | | | | :||| :
0Y 305 -----KAREIDESLIFYKKMELEACVDALLATQIDRNALPRPTYEQDYLKHKIDEYP 359
   || | | | | | | | | | | | | | | :||| :
Db 492 DWLALPAREROP--WFKAADLRA--SAALLLEOASLR-----ROLLLAODELKRLY- 539
   || | | | | | | | | | | | | | | :||| :
0Y 360 QGYPEVSIQHLGYL 373
   || | | | | | | | | | | | | | | :||| :
Db 540 -----LGHLL 543

```

## RESULT 10



```

0Y 154 A-----PRQRLPALACMGVRSLSLEADYRALGCIACLPGRFVAESNEVLLPRIVS 208
Db 1275 LDDSLDDSRMKIM-----AYMAEADMQNY-----RKPIRGPPMPPTSTM 1315
0Y 209 CP-----GFLDDDOO-----EAAARAALQGGGPGPYGPTSTWSTMALGGLPYVL 253
Db 1316 IPRGMEGISEEERQKIMSVMAAEITDSSSVITTSRQPSRSP---SVARKQP--QLMP-- 1367
0Y 254 GQPIRSIPQGI-----VAAMRQHSRSDPSMRQPERY----- 285
Db 1368 PQQAIPRIIPQGLEGLSDEBRTIMSVMAAEFEESRQVPS--RQPSRSPSVNPOQSFP 1426
0Y 286 -----ILRPFR--REVEKTACPSGKKAREIDSLITKKWELACVDAALLATQMDRV 337
Db 1427 IPSEFPIPPGLEDISEEEROKIMSVMAAEVEESRQSLPQRSRSPSVAMI--QAPAV 1484
0Y 338 NAIPPTYQOL-DVLKHKIDELX-----PQGY---PESVIHLGLYFLKMSPE 380
Db 1485 PIIPSGMEDLPEAEERQKIMSVMAEAEIQNFVPSRSPSVMSQPPVPIH-----GLE 1536
0Y 381 DIRKNNVTSLETLKALLEVD-----KGHEMSQAPRRPLPOVATLIRFYKRGOL 431
Db 1537 DLSEEROKIISVMAEAEIDSAKIIPSRSTSSYMPPLPQMSQPEITTTGLEHSSADMEF 1596
0Y 432 DKDLDLTLTAAVPGYGLCSLSPBELSSVPPSSIAWVRPDLDTCPR--OLD--VLYPKARL 488
Db 1597 GRDSSRSQVPIPG-LEMLSEERQI--MSYMA--HALESIIIPSGHIDQIPISLPRHT 1651
0Y 489 AFQÑ---MNGSEYFVQIOSFLGAGPTEDLKALISOONVSMDLATEFKLTDVLPITVAEV 545
Db 1652 GFKPAGIYNEDLEFETERKOREESPTRESGYATSTSYREELAMGDEERMDGLE-DIIRI 1710
0Y 546 QKLLGPHVEGLKAEERHRRPVDWILRQRODD 576
Db 1711 R-----EGAS---RRDSRDEYVLRREED 1731

```

RESULT 13  
E82662 ABC transporter ATP-binding protein Xf1602 [imported] - *Xylella fastidiosa* (strain 9a5c)  
C:Species: *Xylella fastidiosa*  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: E82662  
R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
A:Reference number: A82515; MUID:20365717  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82662  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-283 <SIM>  
A:Cross-references: GB:A0003987; GB:A0003849; NID:99106634; PIDN:AAF84411.1; GSPDB:GN001  
A:Experimental source: Strain 9a5c  
R:Stimpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Brigues, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H  
as-Neto, E.; Docena, C.; El-Dorcy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohnm  
J.D.; Junguelza, M.L.; Kemper, E.L.; Kitzajma, J.P.; Klieger, J.E.; Kurame, E.E.; Laigra  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, J.M.F.; Martino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; M  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sasanaka  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
A:Genetics:  
A:Gene: XF1602

[illegible]

RESULT 14

T42756

5-oxoprolinase (ATP-hydrolyzing) (EC 3.5.2.9) - rat

N:Alternate names: pyroglutamate (ATP-hydrolyzing)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-May-2000

C:Accession: T42756

R:Ye, G.J.; Breslow, E.; Meister, A.

J: Biol. Chem. 271, 32293-32300, 1996

A:Title: The amino acid sequence of rat kidney 5-oxo-L-prolinase determined by cDNA

A:Reference number: 222263; MUID:97113037

A:Accession: T42756

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1288 <YES>

A:Cross-references: EMBL:U70825; NID:g1732064; PID:g11332065; PIDN:AAC52955.1

A:Experimental source: strain Sprague-Dawley; kidney

C:Superfamily: hypodietical protein YKL215c

C:Keywords: homodimer; hydrolase

Query Match	3.3%	Score 107.5	DB 2	length 1288
Best Local Similarity	20.4%	Pred No. 21		
Matches 140	Conservative 69	Mismatches 224	Indels 253	Gaps 34

  

Qy	19	LLFLFLSLGWHPARTLAGEGTGTESAPLGGVL--TTPHNISSLSRQLLGFPCAEVSGL	75
	11		
Db	265	YIFMR-SDGGLPMPAFSGSAVRLTSGPAGGVGYSATYHLEGGP--YIGF--DMGCT	318
	11		
Qy	76	ST-----RRVELAVALAKNNKJSTEDLRCLAHLSLEPPEDLDLPL-----	118
	11		
Db	319	STDVSRVAGEEFHVEASTA---GVTLDAFOL-----DINTVAAGGSRLEFF	362
	11		
Qy	119	-DLLFLFND--AFSGPOACTRFSSRIYKAVNDLLPRGAPEROLLPALACMGVRSGL	174
	11		
Db	363	RSGLFVVGSGESAGAHGP-ACYRKGGPVTVLDANLV-----LGRLLPASPFC-----	408
	11		
Qy	175	LEAEVVRALGGLACDLPGRFVAESAELLFLVSCSGPLDDQDQAEARALOGGGPPY--	232
	11		
Db	409	-----ITPGEGEOLSPSPASRKALAEVVAEIVMS	436
	11		
Qy	233	---GPPSTSVSTMDALNGLLPLVGLGPIIRSIPOGIIVAMRORSSRDPSMROPERTILR	288
	11		
Db	437	FLTNPCPPASQSLSEVAVGFEVRVNAENACRPI-----RALQARGHDS-----	481
	11		

QY 289 PRFRREVEKTACPSG-----KKAREIDESLIFYKKMELE---ACVDAALLATQMD 335  
 Db 482 -----ANVLACFGAGQGHACALRALGMDTVHHRHSGLLSALGLADLVNHEAOEPC 535  
 QY 336 RVNAIPETEDLDVLRKHLDE----LYPGYPSVIOHGLGYLTKMSPEDIRKWNVTSL 390  
 Db 536 SLSTYPEFPAOLDRLSLREOCVADLQVGFPRSOISTESPLHLRQYGTDC----- 587  
 QY 391 ETIKALLEVDKGHMSPPQARRP-----LPQVATLIDRF-VKRG--- 429  
 Db 568 -----ALWMSAHQHHPATACSPRAGDFGAALFVERKMEFGFIIPEPVVDVVRGTGRSG 643  
 QY 430 -QLDKDT-----LDLTAFY-----PGYL----- 447  
 Db 644 LQLE-DTFKIQTPPHVKEVQCYFEGSYQETFPVYLLELQVGHQLOGPCLLIDNNSTIL 702  
 QY 448 -----CSLSPBELSSVPSSIAVPRQDLDL-CDPRLDVLVYPKARLAFONNNGSEYFVKI 502  
 Db 703 VEPGQAEVETDGTDI-RISVGAEGPSMADTRLDPIQLSIF-----SHREFSI 748  
 QY 503 OSFLGAPTEDELKALSOQNVSMDLATPMKLRDVAVLPTVAEYOKLLGPHVEGLKAEBRH 562  
 Db 749 AEQMG-----RIIQRAIS-----TNKERLDPSG-----ALFGPD-GGLVSNAPH 788  
 QY 563 RPRVDMLIRQR-QDDLDLTLGLGLOGG 587  
 Db 789 IPVHLGAMQETVQFOIQLHGLADLHPG 814

## RESULT 15

S23070 erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea

N:Alternate names: 6-deoxyerythronolide B synthase II

C:Species: Saccharopolyspora erythraea

C:Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text-change 21-Jul-2000

C:Accession: S23070; S22011; S23205

R:Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.

Eur. J. Biochem. 204, 39-49, 1992

A:Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of

A:Reference number: S23070; MUID:92155230

A:Accession: S23070

A:Molecule type: DNA

A:Residues: 1-5573 <BEV1>

A:Cross-references: EMBL:X62569

A:Experimental source: strain NRRL 2338

R:Bevitt, D.J.

submitted to the EMBL Data Library, September 1991

A:Reference number: S22011

A:Accession: S22011

A:Molecule type: DNA

A:Residues: 1-184, 'I', 186-301, 'S', 303-521, 523-658, 'A', 660-993, 1001-1212, 'H', 1214-1392, 13

3479, 'DI', 3480-3572 <BEV2>

A:Cross-references: EMBL:X62569; NID:946977; PIDN:CA44448.1; PID:9581651

R:Caiffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.

FEBS Lett. 304, 225-228, 1992

A:Title: Identification of DEBS 1, DEBS 2 and DEBS 3, the multi-enzyme polypeptides of th

A:Reference number: S23103; MUID:92316235

A:Accession: S23205

A:Molecule type: protein

A:Residues: 2-12, 'XXX' <CAP>

A:Experimental source: strain CA340

C:Genetics:

A:Gene: eryA

A:Start codon: GTG

C:Function:

A:Description: catalyzes the construction of a polyketide chain, which is then cyclised

A:Pathway: erythromycin biosynthesis

C:Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car

ort-chain alcohol dehydrogenase homology

C:Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi

F:52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F:561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

F:1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>  
 F:1404-1475/Domain: acyl carrier protein homology <ACP>  
 F:1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>  
 F:2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>  
 F:2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>  
 F:3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>  
 F:3420-3493/Domain: acyl carrier protein homology <ACP2>  
 F:1439/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 3.38; Score 107.5; DB 2; Length 3573;

Best Local Similarity 20.48; Pred. No. 90;

Matches 147; Conservative 63; Mismatches 225; Indels 285; Gaps 35;

QY 10 WSCGDRPSLFLFLFSLGVWHPARTLAGETGESAFLGGLTTHNNTSSLSPRQLLGPC 69  
 Db 1894 WPAGERP-----RRAGVSFGVSG--TNAHYVEEP----- 1923  
 QY 70 AEVSGLSERVRELAVALAOKNVKLSFEOCLAHRLSEPPEDDALPLDLLFLNPDAP 129  
 Db 1924 AEQEAFTER-GPIPFVLSGRSEAVVAQAARALAEHLDTPE-----LGLTDAM 1972  
 QY 130 SGPOACTFFFSRIRK-----ANVDLPRGAPERRLPALACGVCGLSEADY 180  
 Db 1973 TLANGRAFDVRAAYLVGDDRAGVCAEIDLAEGRSADAVAPVTA-----PKRPV 2023  
 QY 181 RALGLACDLPG--RFVSAEAVLLPRLVSCGPPLDQQAARALQGGCPYPSTW 238  
 Db 2024 LVFPGQAGQWGMARNDLLESSEVFESMSRCBAL-----SPHTDW 2064  
 QY 239 SVSTMALRG-----LFLVGOPITRISIPQIVAAARORSSRDSRWPERTILR 290  
 Db 2065 KL--LDVVRGCGGDPPIRHRVVL--QPVLSIMVSLAEIMRAHG-----VTP- 2107  
 QY 291 FRREVEKTACPSGKKARIDESLIFYKKMELEACVDALLATQMDRVNAIPFTEQDLV 350  
 Db 2108 -----AAVGHSGGTAAAHV-AGALSTLEAAKVAALNSQV----- 2142  
 QY 351 KHKIDELYPGYPSVIOHGLGYLTKMSPEDIRKWNVTSLETLKA----- 395  
 Db 2143 --LRELDQG-----GMVSAGASRDE-----LETVLARMDGRVAVAANGPG 2182  
 QY 396 -----LLEVD-----KGHEMSPO--APR--RPLQVATLIDRFVKGROLDKDTIDT 438  
 Db 2183 TSVVAGPTAEIDDEFPAEAEAREMKPRRIAVYASHSPVARIEDRLA-----ELGT 2234  
 QY 439 LTAFFPYGLCSLSPBELSSVPSSIAVPRQDLDLCPDROLDVLYPKARLAFONNNGSEY 498  
 Db 2235 ITAV-----KGSVPLHS--TVTGEYIDT--SANDAST-----WYRDLRPPVL 2272  
 QY 499 FVKIQSLGAPTEDELKALSOQNVSMDLATPMKLRDVAVLPTVAE----- 544  
 Db 2273 F-----EQAVRELTVQGFQD--TFEVRSPHVLMAVEENAEHGAETVCPT 2317  
 QY 545 -----VQKLGHVGLKAER--HRPV-----RDWLIRORODLD 578  
 Db 2318 LRREGSGPHEFLRMLRAHVHGVGADLRPAVAGRPALPTYPPEHOFMRPRRPADVS 2377  
 QY 579 TLGL-----GLGGIPNGYLVLDLS--VOETLSGTPCLLGGPVLTVLAL 621  
 Db 2378 ALGVRAEHPHLLLAADVDPVGHGAVFTGRSLSTDQPLAEHVVGGRTLV-PGSVLVDLAL 2436

Search completed: December 7, 2001, 10:02:37

Job time: 610 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:47:57 : Search time 23.78 Seconds  
(without alignments)  
968.272 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALGRUDPCWCGDRPGSL.....LLGPPVLVALLASTIA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3261	100.0	628	1 MSLN_HUMAN	Q13421 homo sapien
2	1223	3.8	650	1 APP1_HUMAN	P16163 homo sapien
3	111	3.4	1108	1 CYGE_MOUSE	P52785 mus musculu
4	110.5	3.4	5065	1 EPPL_HUMAN	P58107 homo sapien
5	109.5	3.4	834	1 DPOL_THECA	P80194 thermus aqu
6	108.5	3.3	1398	1 DNA2_SCHPO	Q94922 schizosacch
7	107.5	3.3	476	1 AUPL_HUMAN	Q94678 homo sapien
8	107.5	3.3	1288	1 ERY2_SACER	P97608 ratu
9	107.5	3.3	3567	1 ERY2_SACER	Q03132 saccharopol
10	106.5	3.3	653	1 APPI_MOUSE	Q03157 mus musculu
11	106	3.3	690	1 HRPO_BURSO	P35656 burkholderi
12	105	3.2	693	1 PKNL_MYXXA	P33972 myxococcus
13	104.5	3.2	834	1 DPOL_THECA	P52028 thermus aqu
14	104	3.2	1003	1 SYG_CHLTR	Q46371 chlamydia t
15	104	3.2	1142	1 SPK_HUMAN	Q92797 homo sapien
16	103.5	3.2	4128	1 PRKD_MOUSE	P97313 mus musculu
17	102	3.1	1159	1 DP3A_HAEIN	P43743 haemophilus
18	101	3.1	677	1 LCN_HUMAN	P81274 homo sapien
19	100.5	3.1	1073	1 RAG1_ONCMY	Q91187 oncorhynch
20	99.5	3.1	461	1 NCB1_HUMAN	Q02818 homo sapien
21	99.5	3.1	1807	1 ITB4_RAT	Q46633 ratu
22	98.5	3.0	785	1 SYFB_THETH	P27002 thermus aqu
23	98.5	3.0	847	1 IP3L_HUMAN	P27987 homo sapien
24	98	3.0	917	1 IL6B_MOUSE	Q00560 mus musculu
25	98	3.0	1103	1 CYGD_HUMAN	Q02846 homo sapien
26	98	3.0	1108	1 CYGE_RAT	P51840 ratu
27	97.5	3.0	344	1 VIBY_AGRTS	P07169 agrobacteri
28	97.5	3.0	574	1 MP12_RAT	P48966 ratu
29	97.5	3.0	1762	1 DPOO_HUMAN	Q75417 homo sapien
30	97.5	3.0	2035	1 EVPL_MOUSE	Q04952 mus musculu
31	97	3.0	780	1 SAV_STULAC	Q07590 sulfolobus
32	97	3.0	918	1 IL6B_RAT	P40190 ratu
33	97	3.0	1109	1 CYGD_CANFA	O19179 canis famli

34	96.5	3.0	308	1 GDFF_HUMAN	Q99988 homo sapien
35	96.5	3.0	494	1 GP2A_ECOLI	P25552 escherichia
36	96.5	3.0	984	1 PKL2_HUMAN	O16513 homo sapien
37	96	2.9	367	1 CURD_STRON	Q02587 streptomyce
38	96	2.9	919	1 DNL1_HUMAN	P18858 homo sapien
39	95.5	2.9	513	1 ND11_YEAST	P32340 saccharomyc
40	95.5	2.9	5035	1 RYNR_PIG	P16960 mus scrofa
41	95	2.9	991	1 BMPL_MOUSE	P98063 mus musculu
42	95	2.9	1159	1 DP3A_PASMU	Q9CPK3 pasteurella
43	95	2.9	1164	1 YAV1_XANCV	P14727 xanthomonas
44	95	2.9	1698	1 Y076_HUMAN	Q14999 homo sapien
45	95	2.9	2241	1 TEGU_HCMVA	P16785 human cytom

## ALIGNMENTS

RESULT	1	MSLN_HUMAN	STANDARD:	PRT:	628 AA.
AC	Q13421:	MSLN_HUMAN			
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	MESOTHELIN PRECURSOR (CAK1 ANTIGEN).				
GN	MSLN.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Carcinoma;				
RX	MEDLINE-96133892; PubMed-6552591;				
RA	Chang K., Pastan I.;				
RT	"Molecular cloning of mesothelin, a differentiation antigen present				
RT	on mesothelium, mesotheliomas, and ovarian cancers.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:136-140(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.				
RX	MEDLINE-94200897; PubMed-8150545;				
RA	Chang K., Pastan I.;				
RT	"Molecular cloning and expression of a cDNA encoding a protein				
RT	detected by the K1 antibody from an ovarian carcinoma (OVCA-3) cell				
RT	line.";				
RL	Int. J. Cancer 57:90-97(1994).				
CC	-I- FUNCTION: MAY PLAY A ROLE IN CELLULAR ADHESION. ANTIGENIC PROTEIN				
CC	REACTIVE WITH ANTIBODY K1.				
CC	-I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.				
CC	NOT SHED TO THE SERUM OF CANCER PATIENTS.				
CC	-I- TISSUE SPECIFICITY: EXPRESSED IN MESOTHELIAL CELLS, MESOTHELIOMAS,				
CC	OVARIAN CANCERS, AND SOME SQUAMOUS CELL CARCINOMAS.				
CC	-I- PTM: GLYCOSYLATED.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL: U40434; AAC50348.1; .				
DR	MIM: 601051; .				
KW	Signal: Glycoprotein; Antigen; GPI-anchor.				
FT	SIGNAL	1	31	POTENTIAL.	
FT	PROPEP	32	294	REMOVED IN MATURE FORM (POTENTIAL).	
FT	CHAIN	295	586	MESOTHELIN.	
FT	PROPEP	587	628	REMOVED IN MATURE FORM (POTENTIAL).	
FT	CARBOHYD	55	55	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	386	386	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	494	494	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	521	521	N-LINKED (GLCNAC. . .) (POTENTIAL).	

FT LIPID 586 586 GPI-ANCHOR (POTENTIAL).  
SQ SEQUENCE 628 AA; 69044 MM; 2783ED29C3855FD CRC64;

Query Match 100.0%; Score 3261; DB 1; Length 628;  
Best Local Similarity 100.0%; Pred. No. 1.2e-221;  
Matches 628; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MALQLDPCWCGDRPGSLFLFLSLGWHPARTLAGTGFPSAPLGVLTTPHHNIISSIS 60  
DB 1 MALQLDPCWCGDRPGSLFLFLSLGWHPARTLAGTGFPSAPLGVLTTPHHNIISSIS 60  
OY 61 PROLLGFCFAEVSGLSTERVRELAVALAOKNKVLTSTEDQLCLAHRLSEPPEDLALPLDL 120  
DB 61 PROLLGFCFAEVSGLSTERVRELAVALAOKNKVLTSTEDQLCLAHRLSEPPEDLALPLDL 120  
OY 121 LFLNPDPAFSGPOACTRFRFSRTTKANVDLLPRGAPERQRLPALALACWVSGSLSEADY 180  
DB 121 LFLNPDPAFSGPOACTRFRFSRTTKANVDLLPRGAPERQRLPALALACWVSGSLSEADY 180  
OY 181 RALGGLADDLGGRFYAESAEVLLPRLVSCPGPLDDQQAARALOGGGPPYGPSTMSY 240  
DB 181 RALGGLADDLGGRFYAESAEVLLPRLVSCPGPLDDQQAARALOGGGPPYGPSTMSY 240  
OY 241 STMALRGLPLVGPPIRSIPQGIYAAMRORSRDPSPMRQPERTILPRFRREVEKTAC 300  
DB 241 STMALRGLPLVGPPIRSIPQGIYAAMRORSRDPSPMRQPERTILPRFRREVEKTAC 300  
OY 301 PSKKAREIDSLFYKKWELEACVDALLATQMDRVNAIPPTYQOLDVYKHKLELYPQ 360  
DB 301 PSKKAREIDSLFYKKWELEACVDALLATQMDRVNAIPPTYQOLDVYKHKLELYPQ 360  
OY 361 GYPSVIVHGLFLFKMSPEDIRKKNVTSLETALKALLEDKHMSPOAPRRPLQVATL 420  
DB 361 GYPSVIVHGLFLFKMSPEDIRKKNVTSLETALKALLEDKHMSPOAPRRPLQVATL 420  
OY 421 IDRFVKGQGLDKDTLDLTFLTAEPGYLCSLSEBELSVSPSSIMAVRQDDLTCDPROLD 480  
DB 421 IDRFVKGQGLDKDTLDLTFLTAEPGYLCSLSEBELSVSPSSIMAVRQDDLTCDPROLD 480  
OY 481 VLYPKARLAFQNMNSESFEVFKIQSLGAPTEDLKALSOQNVSMPLAFMKRTAVLPL 540  
DB 481 VLYPKARLAFQNMNSESFEVFKIQSLGAPTEDLKALSOQNVSMPLAFMKRTAVLPL 540  
OY 541 TVAEVQKLLGFHVEGLKAEERHPRVDMILRQDDDLTGLGOGGIPNGLYVLDSLQ 600  
DB 541 TVAEVQKLLGFHVEGLKAEERHPRVDMILRQDDDLTGLGOGGIPNGLYVLDSLQ 600  
OY 601 ETLGTPCLLGGPVLVALLASTLA 628  
DB 601 ETLGTPCLLGGPVLVALLASTLA 628

RESULT 2

APPL\_HUMAN STANDARD; PRT; 650 AA.  
ID APPL\_HUMAN STANDARD; PRT; 650 AA.  
AC P51693; O00113;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APPL).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98088960; PubMed=9428684;  
RX Paliaga K., Peraus G., Kieger S., Duvrignat U., Hesse L., Multhaup G.,  
RX Masters C.L., Beyreuther K., Weidemann A.,  
RT "Human amyloid precursor-like protein 1"-cDNA cloning, ectopic  
RT expression in COS-7 cells and identification of soluble forms in the

RT cerebrospinal fluid.";  
RL Eur. J. Biochem. 250:354-363(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98180887; PubMed=9521588;  
RX Leinkert U., Kestila M., Lamerdin J., McCreedy P., Adamson A.,  
RX Olsen A., Trygvason K.;  
RT "Structure of the human amyloid-precursor-like protein gene APPL1 at  
RT 19q13.1";  
RL Hum. Genet. 102:192-196(1998).  
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. C-TERMINALLY  
CC PROCESSED IN THE GOLGI COMPLEX AND IS THEN SECRETED.  
CC -I- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN, PARTICULARLY  
CC IN THE CEREBRAL CORTEX POSTSYNAPTIC DENSITY.  
CC -I- PTM: N- AND O-GLYCOSYLATED.  
CC -I- SIMILARITY: BELONGS TO THE APP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U48437; AAB96331.1;  
CC EMBL: AD000864; AAB50173.1;  
CC MIM: 104775;  
CC InterPro: IPR001868; A4\_APP.  
CC Pfam: PF02177; A4\_EXTRA; 1.  
CC PRINTS: PR00203; AMYLOIDA4.  
CC SMART: SM00006; A4\_EXTRA; 1.  
CC PROSITE: PS00319; A4\_EXTRA; 1.  
CC PROSITE: PS00320; A4\_INTRA; 1.  
CC GlycoProtein: Transmembrane; Signal.  
CC KW SIGNAL 1 21  
CC FT CHAIN 22 650 POTENTIAL.  
CC FT DOMAIN 22 580 AMYLOID-LIKE PROTEIN 1.  
CC FT TRANSMEM 581 603 EXTRACELLULAR (POTENTIAL).  
CC FT DOMAIN 604 650 POTENTIAL.  
CC FT DOMAIN 640 643 CYTOPLASMIC (POTENTIAL).  
CC FT DOMAIN 241 247 CLATHRIN-BINDING (POTENTIAL).  
CC FT DOMAIN 264 268 POLY-GLU.  
CC FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 461 461 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 551 551 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CONFLICT 48 48 P -> A (IN REF. 2).  
SQ SEQUENCE 650 AA; 72202 MM; 121A034B708C67CA CRC64;

Query Match 3.8%; Score 123; DB 1; Length 650;  
Best Local Similarity 19.1%; Pred. No. 0.37;  
Matches 130; Conservative 79; Mismatches 229; Indels 242; Gaps 30;

OY 100 RCLAHRLSEPPEDLALPLDLILF-----LNP 126  
DB 9 RCLSRPQPPPLPL-LPLLLLLRAOPAIGSLAGSPGAPAPSAQVAGLCGLTLHR 67  
OY 127 DAFSG-----PQA-----CTRFPSRTTKANVDLLPRGAPERQRLPALALAC 167  
DB 68 DLRTRWEPPDQRRRCRDPRVLEYCRQMPPELQIARVE-----QATQAIPEMKC 120  
OY 168 WGVRSGLSEADVRALGGLADLPGRFYAESAEVLLP-----RLVSCGPL--DOD 216  
DB 121 GSRSGSCAHPHHQV-FPRC-LRGEFVSEA--LLVPEGCRFLHGRDQCESSRRHOE 176  
OY 217 QOEAAARA---ALGGG-----PPYGPSTWSVSTMDALRGLPLVLGQ 255  
DB 177 AOEACSSOGLILHSGMILPCGSDRFRGVEYCCPPGTPDPSGTA-----VGD 225  
OY 256 PIIRSIPOGIYAAMRORSRDPSPMRQ-ERTILPRFRREVEKTACPSGKKAREIDSLI 314  
DB 226 PSTRSWPGSRVEGADEDEEESFPQVDYFVEPPQAEDEEETVPPSSHTLAV----- 280

```

QY 315 FYKWELEACVADALLATQMDRVNAIPFTYBOLDVLRKHLDELVPQGYEESVIOHGLYLF 374
DQ 281 -----VGKVTPTPTPTGVDIY-----FGMPGEISEHGFRLR 312
QY 375 LKMSPEDIRKNNVTSLETIKALLEVDKHENSPOAPRRRLPQ-----VATLIDRFVKGQ 430
DQ 313 AKMDLEERMRQIN--EYVREMAADNQSKMLPKADQALNHFQSIQTLEEGVSGRQ 370
QY 431 LDKDT-----LDTIATYPCYGLCSLSPBELSSVPPSS---IWAVRP-----468
DQ 371 RLVEHATRVIALINDORRALLEGFLALQAD-----PPOAEVRLALRRYLRAEOKRQ 425
QY 469 -----ODLDTCDP-----ROLDVLYPKARLA-----FOMN 494
DQ 426 HTLRHQVAVAVDEPKAQOMRFQVHTHLQVIEERNQSLGLDQNPILAQELRQIQEL 485
QY 495 GSEYF--VKISFLGAPTEDELKALSQONVMDLATEFMKLRDVAVLPLVAEVOKLQPH 552
DQ 486 HSEHLGPSELEAPAPGSGSEDEKGLQPPDSKD-----TPMTLPKSGTE-QDAASPE 536
QY 553 VEGLK-AEERRRPVADWLRO-----RODDDTLGLGLOGGIPNKYVLVLSVQETL 603
DQ 537 KEKMPLEQYERKNVNASVPKGFPPHSSEIQDELAAPAGTGV-----REAV 582
QY 604 SGPQLLGPFPVTLVALL 623
DQ 583 SGLIMGAGGSLIYLSML 602

RESULT 3
CYGE_MOUSE STANDARD: PRT: 1108 AA.
AC P52785;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANYLYL CYCLASE GC-E PRECURSOR (EC 4.6.1.2) (GUANYLATE CYCLASE 2E).
GN GUCY2E OR GUC2E.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: Liver;
RX MEDLINE=96435433; PubMed=8838319;
RA Yang R.B., Fulle H.J., Garbers D.L.;
RT "Chromosomal localization and genomic organization of genes encoding
RT guanylyl cyclase receptors expressed in olfactory sensory neurons and
RT retina.";
RL Genomics 31:367-372(1996).
CC -1- CATALYTIC ACTIVITY: GMP -> 3',5'-CYCLIC GMP + PYROPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- PTM: THERE ARE 9 CONSERVED CYSTEINE RESIDUES IN SENSORY GUANYLATE
CC CYCLASES, 6 IN THE EXTRACELLULAR DOMAIN, WHICH MAY BE INVOLVED IN
CC INTRA- OR INTERCHAIN DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: LA1933; AAC42081.1; --
CC HSSP: Q02846; 1AWL.
CC MGD: MGI:105123; Gucy2e.
CC InterPro: IPR001828; ANF_receptor.

```

```

DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001054; Guanylyl_cyclase.
DR Pfam: PF01094; ANF_receptor; 1.
DR Pfam: PF00211; guanylate_cyc; 1.
DR SMART: SM00069; pkinase; 1.
DR SMART: SM00044; cycc; 1.
DR PROSITE: PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE: PS01025; GUANYLATE_CYCLASES_2; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Lyase; cGMP synthesis; Signal; Transmembrane; Multigene family;
KW Vision; Glycoprotein.
FT SIGNAL 1 54 POTENTIAL.
FT CHAIN 5 1108 GUANYLYL CYCLASE GC-E.
FT DOMAIN 55 465 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 466 490 POTENTIAL.
FT DOMAIN 491 1108 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 491 811 PROTEIN KINASE-LIKE.
FT DOMAIN 883 1013 GUANYLATE CYCLASE.
FT DISULFD 108 136 BY SIMILARITY.
FT DISULFD 452 452 INTERCHAIN (BY SIMILARITY).
FT DISULFD 460 460 INTERCHAIN (BY SIMILARITY).
FT CARBOHD 300 300 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 1108 AA; 120725 MW; A1CC9E1B1444C803 CRC64;

```

```

Query Match 3.4%; Score 111; DB 1; Length 1108;
Best Local Similarity 19.9%; Pred. No. 5.5;
Matches 163; Conservative 88; Mismatches 261; Indels 306; Gaps 37;

```

```

QY 7 DPCWSGDRPGSLFLFLSLGAV-----HPARTLAGETGTSAPLG-----47
DQ 106 EPCLT-----PGLAVSSALSRVSGLVGPVNPACCPAPALLAQEGVALVPMCGCTRA 161
QY 48 -----GVLTPPHNISSLPQL-----LGFPCAEGVL 75
DQ 162 GTTAPAVTPAADALVYLLRAFRMARVALITAPQDLWVAGRALSTALRARGLPVALVSM 221
QY 76 STE-----RYRE-----LAVALAOKNKLSTEQRLCL-----AHRLEPPEDDALPL 118
DQ 222 ETSDSGAREALGRDRDPVRVYIMVHNSVLGGEEDORYLLEAEELALDGLSVFLPF 281
QY 119 DLLLE-LNPDASFSGPQACTRFF--SRITKANVDLP-----RGAPERQRL 160
DQ 282 DTLHYALSP-----GPEALAAVFNSSQLRRADAVLTLTRCPGGSVODSLRRAQEHGE 336
QY 161 LPAAL-----ACWYRGSLSEADY-----RALGGLACD 189
DQ 337 LPDLNLKQVSPLEGTIVDAVELLAGVKKARITAVGGGWSGASVAVROVREAOVSGF-CG 395
QY 190 LPGR-----FVNAEAEVLLPRIVSCGPPLDDQDQDAARALQGGPR 231
DQ 396 VLGRKEEPSVFLDTDASGEOLFATHLLDPVGLSIRSGTPIH-----PPRGSPA 445
QY 232 YGP-PSTW-----SVSTMDALRGLLPYLQPII-----258
DQ 446 GCPDSCWFDPDVCNGVGEGLVGVFLVIGMLTGAPLAHYLRHLHMQNASGPNK 505
QY 259 -----RSIPQIVAAVRQSSSD-----PSMKPPEPT-----I 286
DQ 506 IILTLEDVTFLHPGSSRKVYQGSRSLSATRSASDIPVS--QPESTWVGLYEGDMV 563
QY 287 LRPRRREVEKTACPSGK-----KAREI-DESLIFPKKWELEACVADALLATQMDRVNAI- 340
DQ 564 WLKFRPGHHNAIRPATYATSKLRELHENVALLYLGLFLGTADSP--ATPGESILAVV 621
QY 341 -----PFTYBOLDVLRKHLDELVPQGYEESVIOHGLYLFKMSPEDIRKNNVTSLET 394
DQ 622 SEHCARGSLHDLAQREIKDMFKSSLLDLIKGMRYLHNRGAHGRKSRNCVVDGRF 681
QY 395 ALLEVDKGHEMSPOAPRRRLPOVAT-----LIDRFVKGQDLKDTL-----436
DQ 682 VLKVTDHGRLLEA-QVLEPPEPSAEDQMTAPPELLRDPSPLEERGTLACGVFSALAIMQ 740

```

```

QY 437 DTAFAYPGYLSLSPSE-----LSSVPP-----SSIMAVRPDDIDT 473
DQ 741 EVCCSTFYAMLELPEVIEQVRSPPLCRPLVSMQDAPMCIMLOCAVHEHLRPS 800
DQ 474 CD-----PROLDVLYPKARLAFQNNNGSEYFKVIOSEFLGAPLEDIKALISQNV 522
DQ 801 MDLAFDLKRSINKRGKRTNIIDSMLEMLRMEYSSNLEDLIR-----ERTELEGEOK 850
QY 523 SMDLATFMKRLTDAVLPPLTVAEVOVKLGPHEGLKAE 560
DQ 851 KTD-----RLTQ-MLPPSVAAEAK-MGTSVEPEYFEE 881

RESULT 4
EPPL_HUMAN STANDARD; PRT: 5065 AA.
ID EPPL_HUMAN
AC P58107;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE EPIPLAKIN (450 KDA EPIDERMAL ANTIGEN).
GN EPIK1 OR EPIPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=21201183; PubMed=11278896;
RA Fujitwara S., Takeo N., Otani Y., Parry D.A.D., Kunimatsu M., Lu R.,
RA Sasaki M., Matsuo N., Khaleduzzaman M., Yoshioaka H.;
RT "Epiplakin, a novel member of the plakin family originally identified
RT as a 450-kDa human epidermal autoantigen: structure and tissue
RT localization."
RL J. Biol. Chem. 276:13340-13347(2001).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN LIVER,
CC SMALL INTESTINE, COLON, SALIVARY GLANDS, STOMACH AND APPENDIX.
CC -1- SIMILARITY: CONTAINS 65 PLECTIN REPEATS.
CC -1- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB051895; BAB40803.1;
KW Coiled coil; Repeat; Structural protein; Cytoskeleton.
FT REPEAT 9 46 PLECTIN 1.
FT REPEAT 47 84 PLECTIN 2.
FT REPEAT 85 122 PLECTIN 3.
FT REPEAT 123 160 PLECTIN 4.
FT REPEAT 166 200 PLECTIN 5.
FT REPEAT 253 290 PLECTIN 6.
FT REPEAT 291 328 PLECTIN 7.
FT REPEAT 330 366 PLECTIN 8.
FT REPEAT 367 404 PLECTIN 9.
FT REPEAT 521 558 PLECTIN 10.
FT REPEAT 578 615 PLECTIN 11.
FT REPEAT 616 653 PLECTIN 12.
FT REPEAT 654 691 PLECTIN 13.
FT REPEAT 692 729 PLECTIN 14.
FT REPEAT 733 767 PLECTIN 15.
FT REPEAT 840 878 PLECTIN 16.
FT REPEAT 898 935 PLECTIN 17.
FT REPEAT 936 973 PLECTIN 18.
FT REPEAT 974 1011 PLECTIN 19.
FT REPEAT 1012 1049 PLECTIN 20.
FT REPEAT 1214 1251 PLECTIN 21.
FT REPEAT 1252 1289 PLECTIN 22.

```

```

FT REPEAT 1290 1327 PLECTIN 23.
FT REPEAT 1328 1365 PLECTIN 24.
FT REPEAT 1366 1403 PLECTIN 25.
FT REPEAT 1539 1576 PLECTIN 26.
FT REPEAT 1577 1614 PLECTIN 27.
FT REPEAT 1615 1652 PLECTIN 28.
FT REPEAT 1653 1690 PLECTIN 29.
FT REPEAT 1694 1728 PLECTIN 30.
FT REPEAT 1865 1902 PLECTIN 31.
FT REPEAT 1903 1940 PLECTIN 32.
FT REPEAT 1941 1978 PLECTIN 33.
FT REPEAT 1979 2016 PLECTIN 34.
FT REPEAT 2017 2054 PLECTIN 35.
FT REPEAT 2192 2234 PLECTIN 36.
FT REPEAT 2235 2272 PLECTIN 37.
FT REPEAT 2273 2310 PLECTIN 38.
FT REPEAT 2311 2348 PLECTIN 39.
FT REPEAT 2352 2386 PLECTIN 40.
FT REPEAT 2726 2768 PLECTIN 41.
FT REPEAT 2769 2806 PLECTIN 42.
FT REPEAT 2807 2844 PLECTIN 43.
FT REPEAT 2845 2882 PLECTIN 44.
FT REPEAT 2886 2920 PLECTIN 45.
FT REPEAT 3260 3302 PLECTIN 46.
FT REPEAT 3303 3340 PLECTIN 47.
FT REPEAT 3341 3378 PLECTIN 48.
FT REPEAT 3379 3416 PLECTIN 49.
FT REPEAT 3420 3454 PLECTIN 50.
FT REPEAT 3794 3836 PLECTIN 51.
FT REPEAT 3837 3874 PLECTIN 52.
FT REPEAT 3875 3912 PLECTIN 53.
FT REPEAT 3913 3950 PLECTIN 54.
FT REPEAT 3954 3988 PLECTIN 55.
FT REPEAT 4328 4370 PLECTIN 56.
FT REPEAT 4371 4408 PLECTIN 57.
FT REPEAT 4409 4446 PLECTIN 58.
FT REPEAT 4447 4484 PLECTIN 59.
FT REPEAT 4488 4522 PLECTIN 60.
FT REPEAT 4862 4904 PLECTIN 61.
FT REPEAT 4905 4942 PLECTIN 62.
FT REPEAT 4943 4980 PLECTIN 63.
FT REPEAT 4981 5018 PLECTIN 64.
FT REPEAT 5022 5056 PLECTIN 65.
FT REPEAT 1935 1951 COILED COIL (POTENTIAL).
FT DOMAIN 2593 2624 COILED COIL (POTENTIAL).
FT DOMAIN 3127 3158 COILED COIL (POTENTIAL).
FT DOMAIN 3651 3692 COILED COIL (POTENTIAL).
FT DOMAIN 4195 4226 COILED COIL (POTENTIAL).
FT DOMAIN 4729 4760 COILED COIL (POTENTIAL).
SQ SEQUENCE 5065 AA; 553078 MW; 378A472DEDA19489 CRC64;

Query Match 3.4%; Score 110.5; DB 1; Length 5065;
Best Local Similarity 21.9%; Pred. No. 49;
Matches 150; Conservative 80; Mismatches 251; Indels 203; Gaps 36;

QY 28 WHPARTIAGETGTSADLGVL--TTPHNISLSLPROLLGFCAEVSGLSTERYELAV 85
DQ 1008 WEQARLLEAVAT-----GGIIDPISHHH-----LMPVAIQGY-VDOQMELAL 1052
QY 86 ALAOKNVKLSLEQRLCAHRLSEP-----PEDIDALPLDLLFLNDAFSG 131
DQ 1053 SSSSEFPPTPGOGRTSYAOLLEBCPRDETSGILLPLPESAPALPTREQVQRSIQAVPG 1112
QY 132 PQACRREFRSRTKAN-VOLLPRG---APEQRLLPALAC---WGVSGSLSEADVAL 183
DQ 1113 AKDGTSLMDLSSCHFTTEQRRGLIEDVQEGRTYVQQLASVQRVQETKLLAQARVWP 1172
QY 184 GGIACDLP-----GRVASEAEVLLPRLVSCPGPLDDODOEARALOGGAPPYG--- 233
DQ 1173 GPRG-EYVAWMLDAGITTOETLELAAGTOS---PAQVADQPAVKACMLKMGTCAGVLL 1228
QY 234 PPSTWVSITMDALR-GLLPV-LGQPIIRS-IPQGIWAAMRORSSSDPSWROPERTILPR 290

```

```

Db 1229 QPSGAKASIAQAVNDGLPTGLGQRLLEAGVAGSGLV-----DPLNNG----- 1271
Qy 291 FRRREVENTACSSGKKAREIDESLIFYKKWELEACVDALLATQMDRVNAIFTEQDVL 350
Db 1272 -RLSVE-DAAKVGVLGRLESQL-----GQERAAA----- 1300
Qy 351 KHKLEDELPQGPSPESVIOHLGYLFLKMSPEDIRKMNVTSEETLKALLEVDKGHEMSQAP 410
Db 1301 -----GYPD-----PYSRASLSLMQAM---ENG-LVQONE 1326
Qy 411 RRLPQV--AT-LIDREVG-----RGOLDKDTLDTLA-----FYPGYLC 448
Db 1327 GLPLQLQVLAIGVYVD--VHGVLHPQAAACRLGDLDTQSVLTAVDKDNKFFDPDSARD 1385
Qy 449 SLSPSEELSS-----VPPSSIMAVRPOLDICDPQDLQVLYPKARLAFQNMGS 496
Db 1386 QVTYQQLRERCVCSSEFGLLLPLPSDTVLEVDHTAVALAMKV--PVSTGRGCSVS 1443
Qy 497 EYFVKIOSFLGAPTEDEKALSQONVMDLATFMKLRITDAVLPITVAEVQRLGPHVEGL 556
Db 1444 LMDLLSEYVQADKRRELVALCRSG---RAAALRQVVSATVALEAERQPLQATFRGL 1499
Qy 557 KAERHRPV--RDMILRQRODL-----DTLGLOGGIPNGYL--VLDLSV 599
Db 1500 RKQVSARDLFRALISRTTLELSCGTTTVKEVAMDSVKRSLESG---NEIAGVLLQGT 1556
Qy 600 QETLSGTPCL---LGSPVLTVL 619
Db 1557 QERMSIPEALRRHILRPGTALVLL 1580

```

RESULT 5  
DPO1\_THECA STANDARD; PRT: 834 AA.

```

ID DPO1_THECA STANDARD; PRT: 834 AA.
AC P80194;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE I, THERMOSTABLE (EC 2.7.7.7) (TAC POLYMERASE 1).
GN POLA.
OS Thermus aquaticus (subsp. caldophilus).
OX Bacteria; Thermus/Delnooccus group; Thermus group; Thermus.
OC NCBI_TaxID=272;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GX24;
RA Kwon S.-T., Kim J.S., Park J.H., Kim H., Lee D.-S.;
RL Submitted (Aug-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-21.
RC STRAIN=GX24;
RA MEDLINE=93285135; PubMed=8508785;
RA Park J.H., Kim J.S., Kwon S.-T., Lee D.-S.;
RT "Purification and characterization of Thermus caldophilus GX24 DNA
RT polymerase."
RL Eur. J. Biochem. 214:135-140(1993).
CC -1- FUNCTION: THERMOSTABLE ENZYME THAT HAS 5'-TO-3' EXONUCLEASE
CC ACTIVITY AND NO 3'-TO-5' EXONUCLEASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC -1- N PYROPHOSPHATE + DNA(N).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
CC -1-
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U62584; AAB81398.1; -
CC PIR; S33287; S33287.
DR

```

```

DR HSSP; P19821; 17AU.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR002298; DNA_POLI.
DR InterPro; IPR001098; DNA_POL_A.
DR InterPro; IPR000513; Exo_N_I.
DR InterPro; IPR003583; HNH_1.
DR InterPro; IPR003584; HNH_2.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF00476; DNA_POL_A; 1.
DR PRINTS; PR00868; DNAPOLI.
DR SMART; SM00475; 53EXOC; 1.
DR SMART; SM00278; Hnh1; 1.
DR SMART; SM00279; Hnh2; 1.
DR SMART; SM00482; POLAC; 1.
DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
DR Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW DNA-binding.
KW DOMAIN.
FT SEQUENCE 834 AA; 93798 MW; A851FF3C3076348E CRC64;
SQ

```

Query Match 3.48; Score 109.5; DB 1; Length 834;  
Best Local Similarity 19.48; Pred. No. 4.7; Mismatches 241; Indels 273; Gaps 34;  
Matches 143; Conservative 80;

```

Qy 48 GLTTPHNISLSLSPRO-----LIGFPCAEEVSGLSTEREVELAVALAOKNKLSTEQ 98
Db 85 GRAPTEDEF---PRQALIKEIVDLGFTRLVPGYADAV---LATLAKNEKEGE 136
Qy 99 LRCLAHRLSEPPEDIALPLDLFLNPDAF-----SGPQACTFFSRTIKAN 146
Db 137 VRILT---ADRLDQLVSDRVAVLHPEGLHITPEWLMOKYGLKEQWDFRALVGDPS 191
Qy 147 VDLPL--RGAPERQRLPALACWCVRGSLSEAD-----VRA-----LG 184
Db 192 -DNLPGVKIGETAL--KLLEKWSLENLKLNDRVKNPVNREKIKHLEDRLSELS 248
Qy 185 GLACDLP-----GREY-AESAIVLPR-----LVSCGPDPDQOQEAARAAL 225
Db 249 RVRDLDPLREVLAQGRERDREGCLAFLERLFEGSLHFEGLLEAPAPLEAPWP----- 302
Qy 226 QGGGPPVY-----PPTWS-----VSTMDLRGL----- 250
Db 303 ---PPEGAFVGFVLSRPEPMMAELKALACRQGVHRAADPLAGLKDKLEVRGLAADL 358
Qy 251 -----PVLGQPII---RSIQGIYAARQSSRSDPSRQRERTILRRF 291
Db 359 AVLASREGDLVPGDDPMLLAVLLDPSWTPPEGVARRRGGEWTEDA---HRLLSERL 414
Qy 292 RR-----EVEK--TACPSGKKAREIDESLIFYKKWELEACVDALLAT 332
Db 415 HRNLKRLQGEKLLMLYHEVEKPLSRVLAHMEAGVADVAIYDALSLAEELRLEE 474
Qy 333 QMDRVNAIFTEYEQDLVLEKHL-DEL-----YQGGP--ESV 366
Db 475 EVFRLAGHPFLNRSRQDERLVLFDELRLPALGKTOKTGKRSSTAVALBALBEAHPIVEKI 534
Qy 367 IOHLGYLFLKMSPEDIRKMNVTSEETLKALLEVDKGHEMSQAPRRPLPQVATLIDREV 426
Db 535 LQH-----RELTKLNTKYVDPLPSLVH-----PNTGRLLTRF-- 566
Qy 427 GRGOLDKDTLDTLTAIFYGYICLSPEELSSVP-----PSSIWAVRPOLD 472
Db 567 -----NQRTATACGRSSSDP--NLQNIPTVTPPLGQRIIRAFVADGALVALDYS 614
Qy 473 TCDPRQDLVLYPKARLAFQNNNGSEYEVKIOSFLGAPTEDEKALSQONVMDLATFEMKL 532
Db 615 QIELFLAHLGSDENLIVFQEGKDIHTQTASMGVPEAVDPLMR--AAKTVNFGVL 672
Qy 533 RTDAVLPITVAEVQKLGLPHVEGLKAERH---RPVMDILRQRODDLDTLGLGLOGGI 588
Db 673 YGMSAHRLS---DELAIPEEAAVAIFERYFQSPKRVAMIEKT-----LEEGR 717

```

QY 589 PNGYLVDLSVQETLSG 605  
 Db 718 KRGYV-----ETFLG 727

RESULT 6  
 ID DNA2\_SCHPO STANDARD; PRT: 1398 AA.

AC Q9URU2; Q9URU8; Q74241;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DNA REPLICATION HELICASE DNA2.  
 GN DNA2 OR SPBC16D10.04C.  
 OS Schizosaccharomyces pombe (Fission Yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetaceae; Schizosaccharomycetaceae;  
 CC Schizosaccharomycetes.  
 CX NCBI\_TaxID=4896;  
 RN (1)  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RX MEDLINE-20341310; PubMed-10880469;  
 RA Kang H.-Y., Choi E., Bae S.-H., Lee K.-H., Gim B.-S., Kim H.-D.,  
 RA Park C., Macneil S.A., Seo Y.-S.;  
 RT "Genetic analyses of Schizosaccharomyces pombe dna2+ reveal that  
 RT dna2 plays an essential role in Okazaki fragment metabolism.";  
 RL Genetics 155:1055-1067(2000).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC SPRAIN-972;  
 RA Wood V., Rajandream M.A., Barrell B.G., Brown D., Churcher C.M.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE OF 1054-1244 FROM N.A.  
 RA Pasion S.G., Forsburg S.L.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REQUIRED FOR DNA REPLICATION. HAS 3' TO 5' ATP-DEPENDENT  
 CC DNA HELICASE ACTIVITY SPECIFIC FOR FORKED SUBSTRATES. HAS A DIRECT  
 CC ROLE CO-ORDINATING THE MULTIENTZYME PROCESS OF OKAZAKI FRAGMENT  
 CC ELONGATION AND MATURATION.  
 CC -1- SUBUNIT: INTERACTS WITH CDC1, CDC24 AND RAD2.  
 CC -1- SIMILARITY: BELONGS TO THE DNA2/NAM7 HELICASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: AF144384; AAD38528.1; -  
 DR EMBL: AL035637; CAB38508.1; -  
 DR EMBL: AF075169; AAC39502.1; -  
 KW Helicase; DNA-binding; ATP-binding; DNA replication.  
 FT NP\_BIND 955 962 ATP (BY SIMILARITY).  
 FT CONFLICT 120 120 MISSING (IN REF. 1).  
 SO SEQUENCE 1398 AA; 157786 MW; C848FA378BAD9D01 CRC64;

Query Match 3.3%; Score 108.5; DB 1; Length 1398;  
 Best Local Similarity 21.3%; Pred. No. 11;  
 Matches 83; Conservative 58; Mismatches 140; Indels 109; Gaps 19;

QY 307 REI-DESLIEFKKWELEVCVDAALATQMDRVNAIPFYEDL----- 348  
 Db 715 REVDDEDEFFKKWEKLLNQERILL--LKRGDVLFPTLEELAVGKTLPLPYTKREDIV 772

QY 349 -----VLKRLDELXPQGPESVIOHLGY-----LFLKSPEDIRKNVNTS--LETL 393  
 Db 773 CLEIDDRFHKFAFLNDNGYPRNPL-HSGSVGERVVIS---DEHGWMSLAKGIVIAIQ 828  
 QY 394 KALLEYDKGHEMSQAPRRPLPOVATLIDRFVKG-----RGOLDKDTLDT 438

Db 829 DSCIEVTRRHRL--HIPMLKMPNDFKKKQVFFGVNEDSKLSIFGSNNHTRRIDXDEPSS 886  
 QY 439 LFAFYPGYLCISPEELSSVPPSSIMAVR-----PQDDTCC--DRQDVLVYPKARL 488  
 Db 887 GIASIRGTL-----MSSVLPDAPLIRDMIRLKPCKFPCNSALIDPEFKCLNEDQIT 939  
 QY 489 AFQNNNGSEYFVKIQSFSGAPTEDELKALSOONVSM-----LATPMKLRTPAVLPLV-- 542  
 Db 940 ALKCHNAEHSYLLGMGCTGKTTTSSLRSLAKKKKILLTSTHLAVDNILKLLKGC 999  
 QY 543 -AEVOKLLGPVHGKAEERHPRVDMILRORQ--DDIDL-----GLGLQG 586  
 Db 1000 DSTIVRLGSPH-----KHPVKEFCLTEGTFEDDLASLKHFEYEDPQIVACSSSLGYH 1052  
 QY 587 GIPN-----GYLVLDLSVQETLSGTPCLLGP 612  
 Db 1053 STEFKRFPDYCIIDEASQIPL--PICLCP 1079

## RESULT 7

ID AUPL\_HUMAN STANDARD; PRT: 476 AA.  
 AC Q9Y679; Q9UN06; Q9Y685;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ANCIENT UBQUITOUS PROTEIN 1 PRECURSOR.  
 GN AUPL.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).  
 RC TISSUE=Plutitary;  
 RA Peng Y., Song H., Dai M., Huang Q., Mao Y., Zhang Q., Mao M., Fu G.,  
 RA Luo M., Chen J., Hu R.;  
 RT "Human ancient ubiquitous protein AUPL isoform gene";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC TISSUE=Hematopoietic;  
 RA Zhang Q., Ye M., Fu G., Zhou J., Guan Z., Huang Q., Xu S., Gu B.,  
 RA Chen S., Shen Y., Chen Z.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC TISSUE=Hematopoietic;  
 RA Gu J., Huang Q., Yu Y., Xu S., Han Z., Fu G., Zhou J., Wang Y.,  
 RA Huang C., Ren S., Yu Y., Chen Z.;

RT "Novel genes expressed in hematopoietic stem/progenitor cells from  
 RT myelodysplastic syndromes patient";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBQUITOUS.  
 CC -1- SIMILARITY: BELONGS TO THE AUPL FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: AF100754; AAD43018.1; -  
 DR EMBL: AF100753; AAD43017.1; -  
 DR EMBL: AF165515; AAF86645.1; -  
 DR EMBL: AF100746; AAD43010.1; -  
 DR MIM: 602434; -  
 DR InterPro: IPR003892; CUE.

DR SMART; SM00546; CUE; 1.  
 KW Signal: Alternative splicing.  
 FT CHAIN 1 37 POTENTIAL  
 FT SIGNAL 38 476 ANCIENT UBIQUITOUS PROTEIN 1.  
 FT DOMAIN 160 165 POLY-PHE.  
 FT VARSPLIT 114 179 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 354 354 S -> P (IN REF. 3).  
 FT SEQUENCE 476 AA; 53028 MW; A6AF3AD84AE86CB CRC64;

Query Match 3.3%; Score 107.5; DB 1; Length 476;  
 Best Local Similarity 22.6%; Pred. No. 3;  
 Matches 114; Conservative 48; Mismatches 157; Indels 185; Gaps 31;

QY 52 TP--HNSSLSPROLGFPCAEVSGLSERRELAVAL--AQKNVKISTQKLCIAHRL 106  
 DB 98 TPFHDNIVMLT-----TCSTVS---ESEAESATGRFPAQAKAPISP-----LAERF 142  
 QY 107 SEPPEDDALPLDLLFLNPDAFSGPACTFFSRITRKANDLLPRGAPERQRLT--PAA 164  
 DB 143 ----EDTEALPLTPLY-----PTCOFF--FTPLNFIPLAFSSPSGQPLINSPS 187  
 QY 165 LACGVRGSLISEADVRLGGLACDPCRFVAESAENVLLPRLVSCPELDDQGEAARAA 224  
 DB 188 FVCMG-RGFM---EMNGRGELVESLK-RFCASSTR--LP-----PTPLLFPEBEATING 233  
 QY 225 LOGGPPRPSTWSVTMDALRGILPVLGPIIRSTIPQGIYAAMROSSRDPSKROPER 284  
 DB 234 REG---LLRFSSWPFSTIQDV-----QPLTLQVGRPLVSV---TVSDASW----- 272  
 QY 285 TILPRFRREVEKTACPSGKKAREIDESLIIFYKKWELEACVDAAALLATQMRVNAIPETY 344  
 DB 273 -----VSELL-----WSL-----FVPEFTV 286  
 QY 345 EQLDVLK--HK---LDLRYOGYPRESYIOHGLYFLKMSPEIDIRKMWNTSLKALLE 398  
 DB 287 YQVRLRLRVHROLGANEFEALRVQOLVAKELGOTGRTLPAD-----KAENHKKR--- 336  
 QY 399 VDKGEMSPQAPRRRLP-----QVATL-----IDRFYKRGQLD-- 432  
 DB 337 -QRHRLRPPQASQSEPPSPGSPDVOLATLAQRVKEVLPVPLVGIORDLAKTCVOLT 395  
 QY 433 -KDTLDTLTAAYPGYL---SLSEELSSVPS-----SIMVRPDDLT 473  
 DB 396 ITNLLEGAVAMPEDITGTQSLPTASASKPSSSGPVTPPTALTFASMA-RQESLD- 453  
 QY 474 CDPROLDVLPKARLAFQNMNGSE 497  
 DB 454 --ERKQALYEYARRRRFERRAOE 474

RESULT 8  
 OPLA\_RAT STANDARD; PRT; 1288 AA.  
 AC P97608;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE 5-OXOPROLINASE (EC 3.5.2.9) (5-OXO-L-PROLINASE) (PYROGLUTAMASE)  
 DE (5-OXOPHASE).  
 GN OPLAH.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Kidney;  
 RX MEDLINE-97113037; PubMed-8943290;  
 RA Ye G.-J., Breslow E., Meister A.;  
 "The amino acid sequence of rat kidney 5-oxo-L-prolinase determined  
 by cDNA cloning.";  
 RL J. Biol. Chem. 271:32293-32300(1996).

RN [2]  
 RP ERRATUM.  
 RA Ye G.-J., Breslow E., Meister A.;  
 RL J. Biol. Chem. 272:4646-4646(1997).  
 CC -1- FUNCTION: CATALYZES THE CLEAVAGE OF 5-OXO-L-PROLINE TO FORM L-  
 CC GLUTAMATE COUPLED TO THE HYDROLYSIS OF ATP TO ADP AND INORGANIC  
 CC PHOSPHATE.  
 CC -1- CATALYTIC ACTIVITY: ATP + 5-OXO-L-PROLINE + 2 H(2)O = ADP +  
 CC PHOSPHATE + L-GLUTAMATE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- TISSUE SPECIFICITY: WELL EXPRESSED IN TESTIS, KIDNEY AND LIVER.  
 CC -1- SIMILARITY: BELONGS TO THE OXOPROLINASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U70825; AAC52955.1; -  
 DR InterPro; IPR002821; Hydatoinase\_A.  
 DR InterPro; IPR003692; Hydatoinase\_B.  
 DR Pfam; PF01968; Hydatoinase\_A; 1.  
 DR Pfam; PF02538; Hydatoinase\_B; 1.  
 KW Hydatoinase.  
 SQ SEQUENCE 1288 AA; 137746 MW; 4530D0781E10C7AD CRC64;

Query Match 3.3%; Score 107.5; DB 1; Length 1288;  
 Best Local Similarity 20.4%; Pred. No. 12;  
 Matches 140; Conservative 69; Mismatches 224; Indels 253; Gaps 34;

QY 19 LFLPLSLGWVHAKRLAGETGTSAPLGVL---TPPHNSSLSPROLGFPAAEVSGL 75  
 DB 265 VLFMR-SDGGLAPMDAFSGRAVLSGPAVGVSATYHLEGCP--VIGF--DMGT 318  
 QY 76 ST-----ERYRELAVALAQKNVKSTQOLRCIAHRLSEPPEDDALPL- 118  
 DB 319 STDVSRVAGEFEHFEVSTA---GVTLQAPOL-----DITVAAGGSRLFF 362  
 QY 119 -DLLFLNPD---AFSGPACTFFSRITKANVDLLPRGAPERQRLT-PALACGVRGSL 174  
 DB 363 RSGLFVGPESAGAHGP-ACYRKGGPVTYDANLV-----LGRLLPASPFC----- 408  
 QY 175 LSEADVRLGGLACDPCRFVAESAENVLLPRLVSCPELDDQGEAARAAALQGGPP- 232  
 DB 409 -----IFGPGEDQPLSEASRKALEAVAMEVNS 436  
 QY 233 ----GPPSTWSVTMDALRGILPVLGPIIRSTIPQGIYAAMROSSRDPSKROPERITLR 288  
 DB 437 FLTNGPCPASQSLSEVAMGFVYANEMCRPT-----RALTOARGHDS----- 481  
 QY 289 PRFRREVEKTACPSG-----KKAREIDESLIIFYKKWELE---ACVDAALLATQMD 335  
 DB 482 -----AHVLACFGAGGQHACATARALGMDTVIHHRHSGLSALGLADLVVHEADQPC 535  
 QY 336 RVNAIPETFEQDLVAKHLDE---LYPGYPRESYIOHGLYFLKMSPEIDIRKMWNTSL 390  
 DB 536 SLSTYPERFAQLDQRLSRLEQCVDALVOGFPSPQISTESTESFLHRYOGTGC----- 587  
 QY 391 ETIKALLEVDGHEMSPPQAPRRP-----LPOVATLIDRE-VKGRG--- 429  
 DB 588 -----ALMVSANQHPRATACSPRAGDGAAFVERKMBEFGFIIPERVYVDVAVKGTGNSG 643  
 QY 430 -QLDKDT-----LDTLTAFY-----PGYL----- 447  
 DB 644 LQLE-DTPKIQTPRHVAKVQCYFEGGYOETPVYLLGELGYGHQLOGPCLIDNNSTIL 702  
 QY 448 ----CSLSEELSSVPSSTIAVNRPDLT--CDPROLDVLPKARLAFQNMNGSEYFVKI 502  
 DB 703 VEPGQAEVTDITGDI-RISVGAEGSPMADTRLDPIQLSIF-----SHREMSI 748

QY 503 OSFLGAPTEDEKALSOONVSMDLATFMKLRDAVLPLTVAEVOQLGPHVEGLKAEEH 562  
DB 749 AEQMG-----RILOFRALIS-----TNKEREDFSC-----ALFGPD-GGLVSNAPH 788  
OY 563 RPYRWMLROR-QDDLDPLGLGCG 587  
DB 789 IPVHLGAMQETVOFOIHLGADLHPG 814

RESULT 9  
ERY2\_SACER STANDARD: PRT: 3567 AA.  
AC 003132; Q54096;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE ERTHRONOLIDE SYNTHASE, MODULES 3 AND 4 (EC 2.3.1.94) (ORF 2) (6-  
DEOXYERTHRONOLIDE B SYNTHASE II) (DEBS 2).  
GN ERYA.  
OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:  
OC Actinomycetales: Pseudonocardiales: Pseudonocardaceae;  
OC Saccharopolyspora.  
OX NCBI\_TaxID=1836;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9120065; PubMed=2024119;  
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;  
RT "Modular organization of genes required for complex polyketide  
RT biosynthesis.";  
RL Science 252:675-679(1991).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL 2338;  
RX MEDLINE=92155230; PubMed=1740151;  
RA Bevilitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;  
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.  
RT Cloning of the structural gene, sequence analysis and inferred domain  
RT structure of the multifunctional enzyme.";  
RL Eur. J. Biochem. 204:39-49(1992).  
CC -1- CATALYTIC ACTIVITY: 6 METHYLMALONYL-COA + PROTONYL-COA -> 7 COA  
CC + 6-DEOXYERTHRONOLIDE B.  
CC -1- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.  
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN  
CC BIOSYNTHESIS.  
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH  
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3  
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH  
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS  
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,  
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,  
CC RESPECTIVELY.  
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES: ACYLTRANSFERASE (AT),  
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACTYL CARRIER  
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),  
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE  
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION  
CC OF THE FULL-LENGTH CHAIN.  
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M63677; AAA26494.1;  
CC EMBL: X62569; CAA44448.1;  
CC HSSP: P28304; 100R.  
DR InterPro: IPR001227; Acyltransf\_domain.

DR InterPro: IPR002085; Adh\_zinc.  
DR InterPro: IPR000794; Ketoacyl-synt.  
DR InterPro: IPR003880; Phosphopantl\_attach.  
DR Pfam: PF00698; Acyl\_transf; 2.  
DR Pfam: PF00107; adh\_zinc; 1.  
DR Pfam: PF00109; ketoacyl-synt; 2.  
DR Pfam: PF00550; pp-binding; 2.  
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; 2.  
DR PROSITE: PS0075; ACP\_DOMAIN; 2.  
DR Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;  
KW Phosphopantetheine; Multifunctional enzyme.  
FT DOMAIN 1 1484  
FT DOMAIN 1485 3567 MODULE 3.  
FT DOMAIN 27 488 MODULE 4.  
FT DOMAIN 559 884 BETA-KETOACYL SYNTHASE 1.  
FT DOMAIN 1130 1301 BETA-KETOACYL REDUCTASE 1 (POSSIBLY  
NON-FUNCTIONAL).  
FT DOMAIN 1397 1467 ACYL CARRIER (ACP) 1.  
FT DOMAIN 1485 1943 BETA-KETOACYL SYNTHASE 2.  
FT DOMAIN 2013 2336 ACYLTRANSFERASE (AT) 2.  
FT DOMAIN 2383 3066 DEHYDRATASE/ENOYLREDUCTASE (DH/ER).  
FT DOMAIN 3139 3322 BETA-KETOACYL REDUCTASE 2.  
FT DOMAIN 3415 3485 ACYL CARRIER (ACP) 2.  
FT ACT\_SITE 202 202 THIOESTER BOND.  
FT ACT\_SITE 1430 1430 ACYL-ENZYME INTERMEDIATE.  
FT BINDING 1651 651 PHOSPHOPANTETHEINE (BY SIMILARITY).  
FT ACT\_SITE 1661 1661 THIOESTER BOND.  
FT ACT\_SITE 2115 2115 ACYL-ENZYME INTERMEDIATE.  
FT NP\_BIND 2961 2978 NADP (ER).  
FT NP\_BIND 3142 3157 NADP (KR).  
FT BINDING 3448 3448 PHOSPHOPANTETHEINE (BY SIMILARITY).  
FT CONFLICT 438 438 R -> A (IN REF. 2).  
FT CONFLICT 480 480 T -> S (IN REF. 2).  
FT CONFLICT 1241 1241 L -> F (IN REF. 2).  
FT CONFLICT 2664 2664 G -> V (IN REF. 2).  
SQ SEQUENCE 3567 AA; 374413 MW; EE6284F4738AA0C0 CRC64;

Query Match 3.3%; Score 107.5; DB 1; Length 3567;  
Best Local Similarity 20.4%; Pred. No. 49;  
Matches 147; Conservative 63; Mismatches 225; Indels 285; Gaps 35;  
QY 10 WSCGDPSSLFLFLSLGCVHPARTLAGEFTGESAPLGLGVITPTTNISLSPLQLGPPC 69  
DB 1885 WPAGERP-----RKAQVSSFCVSG--TNAIVYENP----- 1914  
QY 70 AEVSGLSTERYRELAVALAOKNVKLSTEOQLCLAHRLSEPPEDLDALPLDLLFLNPDAF 129  
DB 1915 AEQEARTRER-GPLPFVLSGRSEAVVAQAARALAEHLRDTPE-----LGLTDAAM 1963  
QY 130 SGPOACTFEFRIRTK-----ANVDLLPRGAPERRLPALACGVCGLSLEADY 180  
DB 1964 TLATGRARFVDRAAVLGDRAQVCCELDALAEGRSADAVAPVISA-----PRRPV 2014  
QY 181 RALGLACDLPg--RFVSAEVLPLRLVSCPGPLDQDOEAARAALOGGPGPYGPSTW 238  
DB 2015 LVFPQGAQWGMANDLLESSEVFASMSRAEVL-----SPHTDW 2055  
QY 239 SVSYMDALRG-----LLPYLGOPITRSTPGIVAAARRSSRSDSPROPERTILPR 290  
DB 2056 KL--LDVVRGGGDPDPRHRVVL-QPYLFSIMVSLAELMRAHG-----VTP- 2098  
QY 291 FRREVEKATSGSKKARIDESTLFYKKWEELACVDALLATODRVNAIPFTYEQDVL 350  
DB 2099 -----AAVYGHSGGITAIAHV-AGALSLDAAKVVALNSQV----- 2133  
QY 351 KHKLDLVPQGVPSVIOHLGYLEFLKMSPEDIRKWNVTSLEFLKA----- 395  
DB 2134 ---LRELDQg-----GMVSGASRDE-----LETVLARMDGRVAVAANGPG 2173  
QY 396 -----LLEVD-----KGHEKSPQ--APR--RPLPOVALILDRFYKGGQLDKDTLDT 438



```

Db 2174 TSVACPTAEIDEEFAAEAREMKPRRIAVRYASHSEPARIEDRLAA-----ELGT 2225
Qy 439 LTATPYGICLSPEELSSVPSSTMAVRQDITCPRQDVLVYPRARLAFQNMNGSEY 498
Db 2226 ITAV-----RGSVPLHS--TVTGEVIDT--SAMDASY-----WYRLRRPVL 2263
Qy 499 FVKIQSLGAPTEDLKALSOQNVSMDLATPMKLRITDAVLEPLTVAE-----544
Db 2264 F-----EQAIVGLVQGF-----TFVEVSPHVLVMAVEETREHGAETCYPT 2308
Qy 545 -----VQKLGPHEVGLKAER-----HRPV-----RDWILRQDLDL 578
Db 2309 LRREOSGHEERLRLRAHVHGVGADLRPAVAGRPALPTPYREHOFNRPHPADVS 2368
Qy 579 TLGT-----GLGGIPNGYLVDLS--VQETLSGPPCLLGGPVLITVAL 621
Db 2369 ALVGRAHEHPLLAAVDVPGHGAFTGRSTDRQPMIAEHVGGRTLV-PSGLVLDLAL 2427

RESULT 10
APPL_MOUSE STANDARD; PRT; 653 AA.
AC 003157;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APLP).
GN APLP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9306322; PubMed=1279693;
RA Masco W., Bupp K., Magendantz M., Gussella J.F., Tanzi R.E.,
RA Solomon F.;
RT "Identification of a mouse brain cDNA that encodes a protein related
RT to the Alzheimer disease-associated amyloid beta protein precursor.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).
CC -1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN. MAY BE PROCESSED
CC IN THE GOLGI COMPLEX; HOWEVER, IT IS NOT YET CLEAR WHETHER APLP
CC IS SECRETED.
CC -1- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; L04538; AAA37247.1; -
CC PIR; A46362; A46362.
CC MGI; MGI:88046; APLP1.
CC InterPro: IPR001868; A4_APP.
CC Pfam; PF02177; A4_EXTRA.1.
CC PRINTS; PR00203; AMYLOIDA.
CC SMART; SM00006; A4_EXTRA.1.
CC PROSITE; PS00319; A4_EXTRA.1.
CC PROSITE; PS00320; A4_INTRA.1.
CC Glycoprotein; Transmembrane; Signal.
KW SIGNAL
FT CHAIN 1 20 AMYLOID-LIKE PROTEIN 1.
FT DOMAIN 21 653 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 584 606 POTENTIAL.
FT DOMAIN 607 653 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 643 646 CLATHRIN-BINDING (POTENTIAL).
FT DOMAIN 263 271 POLY-GLU.
FT CARBOHYD 464 464 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAc. . .) (POTENTIAL).

```

```

SQ SEQUENCE 653 AA; 72751 MW; 56516DC3E4A0E4B0 CRC64;
Query Match 3.38; Score 106.5; DB 1; Length 653;
Best Local Similarity 20.28; Pred. No. 5.4;
Matches 119; Conservative 72; Mismatches 220; Indels 179; Gaps 27;
Qy 135 CTREFSRTKANYDLRPGAPERORLPAALACWVGSGSLSEADVRLAGLADLPGRF 194
Db 94 CRQMPPELHARVE-----QAAQALPMERNCGSTSGRANHNEVY-PFHC-LPGEF 144
Qy 195 VAESAELLP-----RLVSCGPL--DQDQEAARA--ALGGP--PYG-----233
Db 145 VSEA--LLVPEGCRFLHQRMDQCESTRHQEAQCSOGILIHSGMLLPQSDRFR 202
Qy 234 -----PESTNSVSTMDALRGILPVLCOPTIRSIQGIYAAMAROSSRRPSRQF-ER 284
Db 203 GVEYVCCPRPATPNPSSGAA-----GDPSTRSWPLGGAEGGEDEEVESEFPQVDD 254
Qy 285 TILRRFRREVEKTACPSGKKAREIDSLFYKKWLEACVDALALATQMDRNAIPFTY 344
Db 255 YFVER-----FQAEDEEEDEE-----RAPPSSTHTPMVSRVTPTRPT 294
Qy 345 EQLDVLKHKLDLYPGYVESVIGHLYFLKMSPEDIRKNNVTSLETALKALLEVDKGE 404
Db 295 DGVDVY-----FGMPGEIGHEGFLRAKMDLEERRMQIN--EYRMAMADSQSK 343
Qy 405 MSPQAPRRRLPQ-----VATLIDRFYKRGQLDKDT-----LDITATPYGICLSL 451
Db 344 NLPRARQALNEHQSLIQTLEQVSGERQLVEFHATRYALINDORRALEGFALALQ 403
Qy 452 P-----ELSSVPSSTIMAVR-----467
Db 404 GDPQAEVRLALRYLRAEQKQHTLRHQHVAANDPEKAQOMRQVQTNHLYERRM 463
Qy 468 PDLDTCD--PRQDVLVYPRK--LAFQNMNGSEYFVKIQSLGAPTEDLKALSOQNV 523
Db 464 NQSLGLDQNHLAGLELPQIQIELLAEHLGPSE-----LDASVPGSSSEDKSLQPESSK 519
Qy 524 MD-LATFMKLTQ-----AVLRPLVAQVOKLGPHEVGLKAERHNRVVRMLRQ 573
Db 520 DDPVTLPGKSTIDQSSSGSEKELPLPEQYE-QVYNSAPRGFPFHSQ-----IQ 569
Qy 574 QDDLDITGLGGIPNGYLVDLSVQETLSGPPCLLGGPVLITVALLL 623
Db 570 RDELAPSGTGS-----REALSGLLIMGGGSLYLSTLL 605

RESULT 11
HRPO_BURSO STANDARD; PRT; 690 AA.
ID HRPO_BURSO
AC P35656;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPERSENSITIVITY RESPONSE SECRETION PROTEIN HRPO.
GN HRPO.
OS Burkholderia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_Taxid=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RC MEDLINE=93302711; PubMed=8316211.
RA Gough C.L., Genin S., Lopes V., Boucher C.A.;
RT "Homology between the Hrpo protein of Pseudomonas solanacearum and
RT bacterial proteins implicated in a signal peptide-independent
RT secretion mechanism.";
RL Mol. Gen. Genet. 239:378-392(1993).
CC -1- FUNCTION: INVOLVED IN THE SECRETION OF A PROTEINOGENOUS ELICITOR OF
CC THE HYPERSENSITIVITY RESPONSE IN PLANTS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE

```

(POTENTIAL).  
-1 SIMILARITY: BELONGS TO THE FH1PEP (FLAGELLA/HR/INVASION PROTEINS EXPORT PORE) FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, AJ245811; CAB58250.1; -  
DR EMBL, M96533; -; NOT\_ANNOTATED\_CDS.  
DR PIR, S35251; S35251.  
DR InterPro: IPR001712; Bact\_exp\_FH1PEP.  
DR Pfam, PF00771; FH1PEP; 1.  
DR PRINTS; PRO0949; FLGBIOSNFLHA.  
DR PROSITE; PS00994; FH1PEP; 1.  
KW Protein transport; Transport; Inner membrane; Transmembrane;  
KW Hypersensitive response.  
FT TRANSMEM 14 36 POTENTIAL.  
FT TRANSMEM 48 72 POTENTIAL.  
FT TRANSMEM 104 128 POTENTIAL.  
FT TRANSMEM 196 216 POTENTIAL.  
FT TRANSMEM 229 253 POTENTIAL.  
FT TRANSMEM 291 315 POTENTIAL.  
SQ SEQUENCE 690 AA; 73990 MW; DA6FA8F32417E4D2 CRC64;

Query Match 3.3%; Score 106; DB 1; Length 690;  
Best Local Similarity 19.2%; Pred. No. 6.3;  
Matches 73; Conservative 49; Mismatches 116; Indels 142; Gaps 15;

OY 189 DLGRFVA-ESAEVLLPLRLVSGPLDQDOQAARA-ALGGGPRYPGPPSTSVSTMDAL 246  
DB 419 DPGEPVAVPDGHLIP-----DLPEALRAQAVEAAGLNHPAPHIADPAHAQA 467  
OY 247 RGLPLVLCOPITRSIPQGVIAAMRORS-----RDPSPNQPFRTILPRFREVEKTACP 301  
DB 468 DAALSATQGRVERIADHVHVLRSALHFLVGLQETOMLENTVDYGLVAEAK-AVP 526  
OY 302 SGKARKEIDELIFKKWLEKACVDALLATQMDRVNAIPFTYEDLVLKHLDELDPG 361  
DB 527 AORLA-----DVLRLLEQVPIR 545  
OY 362 YPESYIOHLGVLFLKMSPEDIRKMNVTSLFTLKALLEV--DKGHMSFOAPRRPLPOVAT 419  
DB 546 NMRALL-----ESLVVWGPKRKEDTLMVEYVGGDLGRQIAHQA----- 583  
OY 420 LIDRFVKGRLDKDPTLDLTAFYPGYLCSPSELSSVPPSSIAVVRPQ--DLDTCDP 476  
DB 584 -----TGSTRQM-----PAILLDSVEQ-----TVRAKIRPTAGNLTIDP 620  
OY 477 RQLDVLYKKARLAFONMNGSEYFKIQSFLGAPTEDIKALSQONVSMDLATFMKLRIDA 536  
DB 621 QQVEAI-----IMRLRGIMQGNPVETPSALA-----IYTSMDIR--- 654  
OY 537 VLPLVAEVOKLGPVHVEGL 556  
DB 655 -----RYVRRIIEPHIAL 668

RESULT 12  
PKNL\_MYXXA  
ID PKNL\_MYXXA STANDARD: PRT; 693 AA.  
AC P33973;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE SERINE/THREONINE-PROTEIN KINASE PKNI (EC 2.7.1.-).  
GN PKNI.  
OS Myxococcus xanthus.

OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;  
OC Myxococcales; Cyctobacteriineae; Myxococcaceae; Myxococcus.  
OX NCBI\_taxid=34;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-D2F1;  
RX MEDLINE-92069765; PubMed-1835671;  
RA Munoz-Borodo J., Inouye S., Inouye M.;  
RT "A gene encoding a protein serine/threonine kinase is required for  
RT normal development of *M. xanthus*, a gram-negative bacterium."  
RL Cell 67:995-1006(1991).  
CC -1 FUNCTION: PLAYS AN ESSENTIAL ROLE IN PROPER TIMING OF EARLY  
CC DEVELOPMENT EVENTS.  
CC -1 ENZYME REGULATION: MAY BE REGULATED BY CALCIUM OR A CALMODULIN-  
CC LIKE PROTEIN.  
CC -1 DEVELOPMENTAL STAGE: DEVELOPMENTALLY REGULATED TO START  
CC IMMEDIATELY BEFORE SPORE FORMATION.  
CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, M73498; AAA25402.1; -  
DR PIR, A41090; A41090.  
DR HSSP; O63450; 1A06.  
DR InterPro: IPR000719; Euk\_kinase.  
DR InterPro: IPR002290; Ser\_thr\_kin\_acsite.  
DR InterPro: IPR001440; TPR.  
DR Pfam, PF00069; Kinase; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Phosphorylation.  
FT DOMAIN 59 328 PROTEIN KINASE.  
FT NP\_BIND 65 73 ATP (BY SIMILARITY).  
FT BINDING 88 88 ATP (BY SIMILARITY).  
FT ACT\_SITE 180 180 BY SIMILARITY.  
FT MUTAGEN 88 88 K->N: LOSS OF ATP-BINDING.  
SQ SEQUENCE 693 AA; 74173 MW; ED532EABE7215A91 CRC64;

Query Match 3.2%; Score 105; DB 1; Length 693;  
Best Local Similarity 23.8%; Pred. No. 7.5;  
Matches 80; Conservative 37; Mismatches 105; Indels 114; Gaps 19;

OY 157 RORLLPALACGVGSL-----SEADVALG-GIACDLPGRFVAESAVALPR 205  
DB 378 ORRLAPAPAV--PRASLVEVPVQVLRPGSEPVLRSGSGS--RGGLFLHGRVLPPL 432  
OY 206 LVSCP-----GPIDQOQEAARALOGGPPYGPSTWSYSTDA--LRGLPVL-- 253  
DB 433 CSRPLVLELASGPLSV-MCEYVVRVPPAQAQAVMGMPGFGVQVEATVAVLKAAVDALLQ 491  
OY 254 GQPIIRSIPO-----GIVAMRORSRD-----PSNR-- 280  
DB 492 GEP-VRAVPQVPLEDDPAVARLLEAMRORSAGDAVAVALLEPDDSMGTVRLRTREAWMSL 550  
OY 281 -QPERTILPRRREVEKTACPSGKAREIDSL-----IFKKWE-----LEACVDA 328  
DB 551 ESLEHSLTPPQRAQVDALRV---RVREAALGATVQORALYDAMRGNHGVAKCLEAG 606  
OY 329 ILATQMDVNAIPFTYEDLVKHLDELDP--GYPSVIOHLGVL----- 373  
DB 607 L-----TAQDESIRKREFLARPPQAMGTASHQSGALERDGLSOLDPOY 653  
OY 374 --FLKMSPEDIRKMNVTSLFTLKALLEVDKHEMSP 407

Db 654 ERGKLAPLE-----VDMQRYRLRVLGGRATAP 684

RESULT 13

DPOL\_THETH STANDARD; PRT; 834 AA.

AC P52028;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE DNA POLYMERASE I, THERMOSTABLE (EC 2.7.7.7) (TTH POLYMERASE 1).

GN POLA.

OS Thermus aquaticus (subsp. thermophilus).

OC Bacteria; Thermus/Delnococcus group; Thermus group; Thermus.

OX NCBI\_TaxID=274;

RA STRAIN=HB8 / ATCC 27634;

RA Asakura K., Komatsubara H., Soga S., Yomo T., Oka M., Emi S., Urahe I.;

RT "Cloning, nucleotide sequence, and expression in Escherichia coli of DNA polymerase gene (polA) from Thermus thermophilus HB8.";

RL J. Ferment. Bioeng. 76:265-269(1993).

CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE + DNA(N).

CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; D28878; BAA06033.1; -

DR HSSP; P19821; 1TAV.

DR Interpro: IPR002421; 5\_3\_exonuclease.

DR Interpro: IPR002298; DNA\_POLI.

DR Interpro: IPR001098; DNA\_POLI.

DR Interpro: IPR000513; Exo\_N\_I.

DR Interpro: IPR003583; HHH\_1.

DR Interpro: IPR003584; HHH\_2.

DR Pfam; PF01367; 5\_3\_exonuclease; 1.

DR Pfam; PF00476; DNA\_POLA; 1.

DR PRINTS; PRO0868; DNAPOLI.

DR SMART; SM00475; 53EXOC; 1.

DR SMART; SM00278; Hhh1; 1.

DR SMART; SM00279; Hhh2; 1.

DR SMART; SM00482; POLAC; 1.

DR PROSITE; PS00447; DNA\_POLYMERASE\_A; 1.

KW transferase; DNA-directed DNA polymerase; DNA replication; DNA repair; DNA-binding.

KW DOMAIN

FT SEQUENCE 834 AA; 94049 MW; 1A98145DC1A54A9 CRC64;

QY

Query Match 3.28; Score 104.5; DB 1; Length 834;

Best Local Similarity 19.34; Pred. No. 11;

Matches 142; Conservative 81; Mismatches 241; Indels 273; Gaps 34;

Db 48 GVLTPNNISLSPRO-----LLGFCAYSGSLSTEVRRLAVALAKNNKLSREQ 98

Db 85 GRAPTPEDF---PROALAIKELVDLGFTRLEVGYEAD---DVLATLAKAKEGEYEV 137

QY 99 LRCLAHLSEPPEDLDLPLDLFLFNPDAF-----SGPOACTFFFSRTITKAN 146

Db 138 -----HILFADRLDYLDVSRVAVLHPEGHILITPEWLMKEXGLRPEGVDFRALVGPDS 191

QY 147 VDLPL--RGAPEROLLPALACGVKSGLSLEAD-----VRA-----LG 184

Db 192 -DNLPGVKGIGETAL--KLKKEWGSLENLKNDRVKPEVNRREKIKAHLEDLRLSLELS 248

QY 185 GIADLP-----GRPV-AESAELLPR-----LVSCPGPLDOPDOEABARAAL 225

Db 249 RVRTDLPLEVDLACGREPDREGRLAPLEKLEFGSLHFEGLLEAPLEENPMP----- 302

QY 226 QGGPPYG-----PSTWS-----VSTMDLRGL----- 250

Db 303 -----PPEGAFVGVLSRPEPMWELKALACRDGRVHRAADPLAGLDKLEVRGLAKDL 358

QY 251 -----PYLGQPII--RSTFQGIYAAMRQSSSDPSMRQERITLRPF 291

Db 359 AVLASREGLDLPDDEMLLAYLLDPNSTPTPEGVARRRGGEETDAA---HRLLSERL 414

QY 292 RR-----EVEK--TACPSGKKAREIDESLIFYKKELEACVAAALAT 332

Db 415 HRNLKRLGEGERKLMVLHVEKPLSRVLAEMEATGVRRDAVYQALSLAEIRLEE 474

QY 333 QMDRVNAIPETYEQLDVAKHL-DEL-----YPOGYP--ESV 366

Db 475 EVFRLAGHPENLNSRDOLERVLPEDELRLPALGKTOKTGRSTSAVLEALREAHPIYEKI 534

QY 367 IOHLGYLFKMSPPEDIRKMNWTSLETALKALLEVDKGMHMSQARRLPOVATLIDRFVK 426

Db 535 LQH-----RELTKLNTYVDPLPSLVH-----PRGRLHTR-- 566

QY 427 GRGLDRDTLDTLTAFFPGYLSLSPELSSVP-----PSSIMAVRPQDLD 472

Db 567 -----NOTATATGRILSSSDP-NLONIPIVPTLGIORIRAFVAEAGMALVALDYS 614

QY 473 TCDDPRLDVLTKARLAFQNNNGSEFYVKIOSFLGAFTEDLKALSOONVSMIDLATMKL 532

Db 615 QIELRVLAHLSGDENLIRVEQEGKDIHQTSMMGEGVPEAVDLMRR--AAKTVNGVVL 672

QY 533 RTDVAVLPLTVAEVOKLLGPHVEGLKAEERH---RPVRDWLIRQRDLDLTLGLGLOGI 588

Db 673 YGMSAHRLS---OELAIPEEAVALFETRYFQSPFKRAMTEKT-----LEGR 717

QY 589 PNGYLVDLSVOETLSSG 605

Db 718 KRQYV-----ETLFG 727

RESULT 14

SYG\_CHLTR STANDARD; PRT; 1003 AA.

AC 046371; 084802;

DT 01-NOV-1997 (Rel. 35, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GLYCYL-TRNA SYNTHETASE (EC 6.1.1.14) (GLYRS) [INCLUDES: GLYCYL-TRNA DE SYNTHETASE ALPHA CHAIN (GLYCINE--TRNA LIGASE ALPHA CHAIN); GLYCYL-TRNA DE SYNTHETASE BETA CHAIN (GLYCINE--TRNA LIGASE BETA CHAIN)].

GN GLYRS OR GLYS OR GLYO OR CTY96.

OS Chlamydia trachomatis.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=613;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D/UV-3/CX;

RX MEDLINE=99000809; Pubmed=9784136;

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;

RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";

RL Science 282:754-759(1998).

```

CC -1- CATALYTIC ACTIVITY: ATP + L-GLYCINE + TRNA(GLY) = AMP +
CC PYROPHOSPHATE + L-GLYCYL-TRNA(GLY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U20547; AAC62982.1; -
DR EMBL: AEO01352; AAC68391.1; -
DR InterPro: IPR002106; AA_TRNA_Ligase_II.
DR InterPro: IPR002310; tRNA_synth_2e.
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02091; tRNA_synth_2e; 1.
DR Pfam: PF02092; tRNA_synth_2f; 1.
DR PRINTS: PRO1044; TRNASYNTHGA.
DR PRINTS: PRO1045; TRNASYNTHGB.
DR PROSITE: PS00179; AA_TRNA_Ligase_II_1; FALSE_NEG.
DR PROSITE: PS00339; AA_TRNA_Ligase_II_2; 1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KM Complete proteome.
FT DOMAIN 1 310 GLYCYL-TRNA SYNTHETASE ALPHA CHAIN.
FT DOMAIN 10 1003 GLYCYL-TRNA SYNTHETASE BETA CHAIN.
FT VARIANT 186 186 D -> G (IN SEROVAR L2).
FT VARIANT 221 221 F -> S (IN SEROVAR L2).
FT VARIANT 405 405 S -> P (IN SEROVAR L2).
FT VARIANT 424 424 W -> C (IN SEROVAR L2).
FT VARIANT 565 565 I -> V (IN SEROVAR L2).
FT VARIANT 704 704 T -> P (IN SEROVAR L2).
FT VARIANT 738 738 S -> N (IN SEROVAR L2).
FT VARIANT 838 838 A -> E (IN SEROVAR L2).
FT VARIANT 864 864 V -> I (IN SEROVAR L2).
FT VARIANT 1001 1001 G -> D (IN SEROVAR L2).
FT CONFLICT 602 618 ORFPEPNOGGEITNRF -> PYLPTKTYARRNHOSI
FT (IN REF. 1).
FT S -> SARRAS (IN REF. 1).
FT CONFLICT 828 828 MY -> TH (IN REF. 1).
FT CONFLICT 902 903
FT SEQUENCE 1003 AA; 112576 MW; 79EB918FE19D39E9 CRC64;
SQ

```

Query Match 3.28; Score 104; DB 1; Length 1003;  
 Best Local Similarity 21.48; Pred. No. 15;  
 Matches 130; Conservative 72; Mismatches 221; Indels 184; Gaps 34;

```

QY 38 ETGESAPLGGVLTTPHNISLSPPRL---LGFPCAEGVSGLSTERYRELAVALLAQKNVK 93
DB 318 EIGSEELPATEFVPTGIGQOLESIAKKLADGIAVKHLEVLTPT---RRALAC----- 366
QY 94 LSTQQLCLAHRLSEPPEDDLALPLDLLLPLNPAFSGPQACTFFSRITRANVLLPRLG 153
DB 367 -----IGLSHVTIRPSEKKGPLSLFLM---TGGSVSPQG-EOGFFS---SHGISISHRS 415
QY 154 APERQRLPALACAGVRSGLSADVRLAGLACDLPGRFVASAEVL---LRLVSC- 209
DB 416 ALDO-----PSKI--WRVA-----SINGDYLVLVPEE-RKETAAILVELQLINSI 461
QY 210 --PGPLDODQOEARAALOGGPGPYGPSTWSVTMDALRG--LLPVLGPPIRSIPOGI 265
DB 462 RFPQKMMWD-----NGVEYARIRMLV---ALYDQILP-----ISLGF 498
QY 266 VAAAROSSRDPSV--RO-PEPTLRPRFRREV--KTACPSGKKAREIDSLIFYKWE 320
DB 499 V-----SSGWTSGHQLDRQLTIPSSNMKYVDLRSAC-----YIVSQKE 539
QY 321 LEACVDALLATQMDRNAIIP-----FTYQDLVLLKKHLDLVPQGYPE---SVI 367
DB 540 RRAIIKQGLQNLTDQIVALAIPHLIDETVFLTEHPFVISAQFDPACSLPKELLIAEMI 599

```

```

QY 368 OHIGYLFLKMSPEDIRKWNYSLETLEKALLEVD-----KGHEMSPQAPRRRLPQVAT 419
DB 600 QHORFY-----PTIONMGEITN-----RFLIVCDNSPIDSIVEGNEKA----- 637
QY 420 LIDRFVKGROLDKDTLDLTLPATFYPGYLCSPBELSSVPPSS-----IWAVRP-- 468
DB 638 LAPRLTGNGNLFKODLLTPLSSFVEKLSVYTFESLSGLADKTRSLKHLHEAVALLPLC 697
QY 469 --QDLDTCDROLDVLYPKARLAFQNMNGSEYFKTQSPFLGAPTEDLKALSOQNVSMDL 526
DB 698 AKEDIDTA-----IHYCKADLVSSVVA---EPEELQING-----RYIQNASLSR 740
QY 527 ATFW-----KLRTDAVLPLTVAEVOKLGLPHEGLKAEERHRRPVRMILROR 573
DB 741 AALAIIEHLOHTITGSSISTTGALLSLDRIDNLSCLFILGLPTSSHP---YALRRQ 797
QY 574 ODDLDLT 580
DB 798 SLEILTL 804

RESULT 15
SPK_HUMAN STANDARD; PRT; 1142 AA.
AC Q92797; O00733; O00689;
DT 20-AUG-2001 (Ref. 40, Created)
DT 20-AUG-2001 (Ref. 40, Last sequence update)
DE 20-AUG-2001 (Ref. 40, Last annotation update)
GN SPK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND MASS SPECTROMETRY.
RC TISSUE=Colon carcinoma;
RX MEDLINE=96354916; Pubmed=8769423;
RA Keon B.H., Schaefer S., Kuhn C., Grund C., Franke W.W.;
RT "Symplekin, a novel type of tight junction plaque protein.";
RL J. Cell Biol. 134:1003-1018(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=98252805; Pubmed=9585442;
RA Alwazzan M., Hamshire M.G., Lennon G.G., Brook J.D.;
RT "Six transcripts map within 200 kilobases of the myotonic dystrophy
expanded repeat.";
RL Mamm. Genome 9:485-487(1998).
CC -1- FUNCTION: SPECIFIC COMPONENT OF THE TIGHT JUNCTION (TJ) PLAQUE,
CC BUT MIGHT NOT BE AN EXCLUSIVELY JUNCTIONAL COMPONENT. MAY HAVE A
CC HOUSE-KEEPING RULE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF ADHESION PLAQUES (MAJOR)
CC AND NUCLEOPLASM (MINOR) (IN CELLS WITH TJ). NUCLEOPLASM (IN CELLS
CC WITHOUT TJ).
CC -1- TISSUE SPECIFICITY: IN TESTIS, EXPRESSED IN POLAR EPITHELIA AND
CC SERTOLI CELLS BUT NOT IN VASCULAR ENDOTHELIA. THE PROTEIN IS
CC DETECTED IN STOMACH, DUODENUM, PANCREAS, LIVER, FETAL BRAIN,
CC CARCINOMAS, LENS-FORMING CELLS, FIBROBLASTS, LYMPHOCYTES, LYMPHOMA
CC CELLS, ERYTHROLEUKEMIA CELLS BUT NOT IN ENDOTHELIUM OF VESSELS,
CC EPIDERMIS, INTERCALATED DISKS, PURKINJE FIBER CELLS OF THE HEART
CC AND LYMPH NODE.
CC -1- DIFFERENTIAL DIAGNOSIS OF TUMORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC -----  
DR EMBL: U49240: AAC50667.1: -  
DR EMBL: Y10931: CAA71861.1: -  
DR MIM: 602388: -  
KW Cell adhesion; tight junction; Nuclear protein.  
FT DOMAIN 213 228 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 709 712 POLY-SER.  
FT DOMAIN 1036 1043 POLY-SER.  
SQ SEQUENCE 1142 AA: 126499 MW: 46FF1BAC561C4409 CRC64;

Query Match 3.2%; Score 104; DB 1; Length 1142;  
Best Local Similarity 19.2%; Pred. No. 18;  
Matches 122; Conservative 82; Mismatches 231; Indels 202; Gaps 23;

OY 35 LAGETGTE-SAPLGCVLTTPHNISLSRQLGFPQA-----EYSGLSTE-RVRE 82  
DB 257 ISGSDTDITAEFLQPLTPDPAVNLVISMVYLPEAMPASFOAIYTPVESAGTEAQIKH 316  
OY 83 LAVALA-----OKNVLTSTEOALRCLAHRLSEPPEDDALPLDUL 121  
DB 317 LARLMATQMTAAGLCPGYEQTKCKEPEKEKVVYTESYLIRRLSAQQAISV--VSSL 374  
OY 122 LFLNDPAESGPOACTR-----FESRITKANVDLPRGAP 155  
DB 375 SMSPLLEEBAPQAKRRPEPIIPVTPRLAGAGRRKKIFRLSDVLRPLTDQVEAMKLGAV 434  
OY 156 ERQRLPLPALACWG--VRGSLSEADYRALGCLDLPGRVASEAEVLLPRLYSCQCP 212  
DB 435 KRILAEKAVACSGAAQYRIKILASLVTOFNSGLAEVLS-FILEDVRA--RLDLAFAM 490  
OY 213 LDODQGEAARALOGGPPYGPSTWSVSTMALRGLLP-----VLGP 256  
DB 491 LYQETNATLAAAGASGLDKY-----EDCLIRLSGLQEKDPQDKGIFTKVVLEAP 540  
OY 257 IIRSIPOGIVAAWRORS-----RDPSSWRQPERTILRPFRREVEKTACPSGKKAR 307  
DB 541 LITESALEVVRKYCEDESTYLGMSLRLRIKRRSR-----QFOYLHVLDLSSHEKX 595  
OY 308 EIDESLIF---YKKWELEACVD-----AALLATQMDRVNAIPFTYEOL 347  
DB 596 VRSQALLPFRMYEKEQLEVEKFAALNYQLLVHPNPSPVLEGADKDEVAAPMT--- 651  
OY 348 DVLKHKLDLDELVPQYPSVIOHLGYFLKMSPEDIRKMNVTSLFTLKALLEVDKGHEMS 407  
DB 652 -----EETVQCLYLYLALDPQNHKLIELLAAYTTEATADIKRTVLYRI 695  
OY 408 QAPRRPL---POVATLIDRFYKGRGQLDKDTLDTLTAFFPGYLCSLSPPELSSVPPSSI 463  
DB 696 EQPIRGMGWNSPELLLVENCPCGAETLVTRCLHSLT-----DKVPPSPE 740  
OY 464 WAVRPQDDLTCDPRQDDLVLPKARLAFQNMNGSEYFVKIQSFLGAPTEDLKALSOQNV 523  
DB 741 LVRKVRDL-----YHKRLPDVRRLIPVING-----LEKKEVI 772  
OY 524 MDLATEFMKLRTDAVLPVVAEV-OKLIG-PHVEGLKA 558  
DB 773 QALPKLIKLN-----PIYKVEYFNRLLTQHGEGNSA 804

Search completed: December 7, 2001, 09:48:50  
Job time: 53 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2001, 09:48:17 ; Search time 55.89 Seconds  
(without alignments)  
1643.568 Million cell updates/sec

Title: US-09-215-035-2  
Perfect score: 3261  
Sequence: 1 MALGRDPCWSCGRDRGSL.....LLGPPVTLVALLASTLA 628

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: SP\_Archaea:\*  
3: SP\_Bacteria:\*  
4: SP\_Fungi:\*  
5: SP\_Human:\*  
6: SP\_Invertebrate:\*  
7: SP\_Mammal:\*  
8: SP\_mhc:\*  
9: SP\_Organelle:\*  
10: SP\_Phage:\*  
11: SP\_Plant:\*  
12: SP\_Podent:\*  
13: SP\_Virus:\*  
14: SP\_Vertebrate:\*  
SP\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3110	95.4	630	4 Q9BR17	Q9BR17 homo sapien
2	3055	93.7	622	4 Q14859	Q14859 homo sapien
3	3042.5	93.3	621	4 Q9BTR2	Q9BTR2 homo sapien
4	1710	52.4	625	11 Q9ERA7	Q9ERA7 mus musculu
5	1707	52.3	625	11 Q61468	Q61468 mus musculu
6	1526	46.8	698	4 Q9UK57	Q9UK57 homo sapien
7	133	4.1	698	4 Q9UK57	Q9UK57 pseudomonas
8	131.5	4.0	797	4 Q9UGS3	Q9UGS3 homo sapien
9	126	3.9	939	4 Q9WJ0	Q9WJ0 corynebacte
10	125.5	3.8	2109	8 Q32836	Q32836 pelargonium
11	124	3.8	1906	4 Q15029	Q15029 homo sapien
12	124	3.8	2011	4 Q9P0A9	Q9P0A9 homo sapien
13	120	3.7	2218	10 Q9SL26	Q9SL26 arabidopsis
14	117.5	3.6	3574	10 Q9AUB4	Q9AUB4 arabidopsis
15	116.5	3.6	2015	5 Q9U5Y1	Q9U5Y1 dictyosteli
16	116	3.6	1313	4 Q75033	Q75033 homo sapien
17	116	3.6	1429	5 Q9W5A4	Q9W5A4 drosophila
18	115	3.5	4247	2 Q9L8H4	Q9L8H4 streptomyc
19	114.5	3.5	756	2 Q9A888	Q9A888 caulobacter

20	114	3.5	1067	5 Q9Y1R2	Q9Y1R2 drosophila
21	113.5	3.5	817	5 Q9NK95	Q9NK95 drosophila
22	113.5	3.5	1012	4 Q95712	Q95712 homo sapien
23	112.5	3.4	730	6 Q9TUS2	Q9TUS2 equus caball
24	112.5	3.4	1012	4 Q9UKV7	Q9UKV7 homo sapien
25	111.5	3.4	613	12 Q9W8X7	Q9W8X7 porcine ade
26	111.5	3.4	4957	4 Q14687	Q14687 homo sapien
27	111.5	3.4	5262	4 Q14686	Q14686 homo sapien
28	111	3.4	1596	13 Q918E1	Q918E1 fuigu rubrip
29	111	3.4	3851	4 Q43161	Q43161 homo sapien
30	111	3.4	3926	4 Q9UPA5	Q9UPA5 homo sapien
31	110.5	3.4	340	10 Q22706	Q22706 arabidopsis
32	110	3.4	912	2 Q9K3Y2	Q9K3Y2 streptomyc
33	110	3.4	1770	2 Q50857	Q50857 myxococcus
34	109	3.3	1171	2 Q9XAH6	Q9XAH6 streptomyc
35	108.5	3.3	636	11 Q9DOC6	Q9DOC6 mus musculu
36	108.5	3.3	652	2 Q9HTA6	Q9HTA6 pseudomonas
37	108.5	3.3	1398	3 Q9UR02	Q9UR02 schizosacch
38	108.5	3.3	2361	5 Q94226	Q94226 caenorhabdi
39	108	3.3	283	2 Q9PD02	Q9PD02 xyella fas
40	108	3.3	793	4 Q9H5Z2	Q9H5Z2 homo sapien
41	107.5	3.3	757	2 Q9AJN7	Q9AJN7 athrobacte
42	107	3.3	451	2 Q9A844	Q9A844 caulobacter
43	107	3.3	1715	6 Q9GLM4	Q9GLM4 bos laurus
44	106	3.3	391	2 Q92360	Q92360 streptomyc
45	106	3.3	527	2 Q9RWV6	Q9RWV6 delnococcus

## ALIGNMENTS

RESULT 1  
Q9BR17 PRELIMINARY: PRT: 630 AA.  
ID Q9BR17  
AC Q9BR17  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DR 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DN C335H7.1 (MESOTHELIN).  
GN MSIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Codley V.  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL031258; CAC37289.1;  
SQ SEQUENCE 630 AA; 68985 MW; FA17E3609B6CC9CA CRC64;

Query Match	Best Local Similarity	95.4%;	Score 3110;	DB 4;	Length 630;
Matches 608;	Conservative	4;	Mismatches 16;	Indels 2;	Gaps 2;
QY 1	MALGRDPCWSCGRDRP-GSILFLFLFSIGWVHPARTLAGETGESAPLGGVLTTPHNIS	58			
DB 1	MALPAPARLSCGCPALGSLFLFLFSIGWOPSTTLTGEGQENAPLDGVLANPNNIS	60			
QY 59	LSPROLGFPCAEEVSGSLSTERYRELAVALAOKNVKLTSTOLRCNLRSEPPEDDALPL	118			
DB 61	LSPRQLGFPCAEEVSGSLSTERYRELAVALAOKNVKLTSTOLRCNLRSEPPEDDALPL	120			
QY 119	DLLFLNPDATFSGPACRTRFRSRTIKAVNDLLPKAPERORLLPALACWVGKSLSEA	178			
DB 121	DLLFLNPDATFSGPACRTRFRSRTIKAVNDLLPKAPERORLLPALACWVGKSLSEA	180			
QY 179	DYRALGGLACDIPGRFVESAEEVLPLRVSCGPPLDDOQOEAAARALOGGGPPYPTW	238			
DB 181	DYRALGGLACDIPGRFVESAEEVLPLRVSCGPPLDDOQOEAAARALOGGGPPYPTW	240			
QY 239	SVSTMALRGLPLVGLGPIIRISIPQIVAAWKRSSRDPSPWROPERTILRPFRREVEKT	298			

```

|||||
Db 241 SVSTMDALRGLLPVLGQPIIRSIPOGIYAAMRQSSRPSWQRPRTILRPFRREREVET 300
QY 299 ACPGSKKAREIDESLIFKKWELEACVDAALLATQMDRVNAPFYEDLDVKKHLDLXL 358
Db 301 ACPGSKKAREIDESLIFKKWELEACVDAALLATQMDRVNAPFYEDLDVKKHLDLXL 360
QY 359 POGYVESYIOHLGYFLFKMSPEDIRKMNVTSLKETKALLEVDKGHEMSPOAPRRRLPOVA 418
Db 361 POGYVESYIOHLGYFLFKMSPEDIRKMNVTSLKETKALLEVDKGHEMSPOAPRRRLPOVA 420
QY 419 TLIDRFVKGROLDKDTLDTLTAFFPGYLCSLSPBELSSVPPSSIWAVRPODLTCDPRQ 478
Db 421 TLIDRFVKGROLDKDTLDTLTAFFPGYLCSLSPBELSSVPPSSIWAVRPODLTCDPRQ 480
QY 479 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEDLKALSOQNVSMDLATEPKKLTDAVL 538
Db 481 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEDLKALSOQNVSMDLATEPKKLTDAVL 540
QY 539 PLTVAEVOKLLGPHVEGLKAEBHRHVRDWMILRQRDDLDLTGLGLOGGIPNGYLVLDLS 598
Db 541 PLTVAEVOKLLGPHVEGLKAEBHRHVRDWMILRQRDDLDLTGLGLOGGIPNGYLVLDLS 600
QY 599 VOETLSGTPCLLGGPVLVTALLASTLA 628
Db 601 MOEALSGTPCLLGGPVLVTALLASTLA 630

```

```

RESULT 2
ID 014859 PRELIMINARY; PRT; 622 AA.
AC 014859;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PRE-PRO-MEGAKARYOCYTE POTENTIATING FACTOR PRECURSOR.
GN MPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95394969; PubMed=7665620;
RA Kojima T., Oh-peda M., Hattori K., Taniguchi Y., Tamura M., Ochi N.,
RA Yamaguchi N.;
RT "Molecular cloning and expression of megakaryocyte potentiating factor
cDNA.";
RL J. Biol. Chem. 270:21984-21990(1995).
DR EMBL; D49441; BAA08419.1;
DR InterPro; IPR000585; Hemoexin.
DR PROSITE; PS00024; HEMOEXIN; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 286 MEGAKARYOCYTE POTENTIATING FACTOR.
SQ SEQUENCE 622 AA; 68037 MW; 1F0FE52883CA6C13 CRC64;

```

```

Query Match 93.7%; Score 3055; DB 4; Length 622;
Best Local Similarity 95.4%; Pred. No. 4,8e-226;
Matches 601; Conservative 3; Mismatches 16; Indels 10; Gaps 3;
QY 1 MALQRIDP-CWSCGDRP-GSLIFLLFSLGWHPARTLAGETGESAPLGAVLTTPHNISS 58
Db 1 MALPTARILGSCGCPALGSLIFLLFSLGWQPSRTLAGETGEAPLDGYLANPNISS 60
QY 59 LSPRLGFPCEAVESGLSTERYRELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDALPL 118
Db 61 LSPRLGFPCEAVESGLSTERYRELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDALPL 120
QY 119 DLLFLFNDASSGQACRFRFSRITKANVDLLPRGAPRORLLPRAALACWGVRSLSSEA 178
Db 121 DLLFLFNDASSGQACRFRFSRITKANVDLLPRGAPRORLLPRAALACWGVRSLSSEA 180

```

```

QY 179 DVRALGLACDLPGRFVAESAELLRLVSCGPIIDODQOEARAALOGGPPYPSTW 238
Db 181 DVRALGLACDLPGRFVAESAELLRLVSCGPIIDODQOEARAALOGGPPYPSTW 240
QY 239 SVSTMDALRGLLPVLGQPIIRSIPOGIYAAMRQSSRPSWQRPRTILRPFRREREVET 298
Db 241 SVSTMDALRGLLPVLGQPIIRSIPOGIYAAMRQSSRPSWQRPRTILRPFRREREVET 300
QY 299 ACPGSKKAREIDESLIFKKWELEACVDAALLATQMDRVNAPFYEDLDVKKHLDLXL 358
Db 301 ACPGSKKAREIDESLIFKKWELEACVDAALLATQMDRVNAPFYEDLDVKKHLDLXL 360
QY 359 POGYVESYIOHLGYFLFKMSPEDIRKMNVTSLKETKALLEVDKGHEMSPOAPRRRLPOVA 418
Db 361 POGYVESYIOHLGYFLFKMSPEDIRKMNVTSLKETKALLEVDKGHEMSPOAPRRRLPOVA 420
QY 419 TLIDRFVKGROLDKDTLDTLTAFFPGYLCSLSPBELSSVPPSSIWAVRPODLTCDPRQ 478
Db 421 TLIDRFVKGROLDKDTLDTLTAFFPGYLCSLSPBELSSVPPSSIWAVRPODLTCDPRQ 480
QY 479 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEDLKALSOQNVSMDLATEPKKLTDAVL 538
Db 481 LDVLYPKARLAFQNMNGSEYFVKIOSFLGAPTEDLKALSOQNVSMDLATEPKKLTDAVL 540
QY 539 PLTVAEVOKLLGPHVEGLKAEBHRHVRDWMILRQRDDLDLTGLGLOGGIPNGYLVLDLS 598
Db 541 PLTVAEVOKLLGPHVEGLKAEBHRHVRDWMILRQRDDLDLTGLGLOGGIPNGYLVLDLS 600
QY 599 VOETLSGTPCLLGGPVLVTALLASTLA 628
Db 601 MOEALSGTPCLLGGPVLVTALLASTLA 622

```

```

RESULT 3
ID 09BTR2 PRELIMINARY; PRT; 621 AA.
AC 09BTR2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MESOTHELIN.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=PANCREATIC ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003512; AAH03512.1;
SQ SEQUENCE 621 AA; 67908 MW; 1A7367740956997B CRC64;

```

```

Query Match 93.3%; Score 3042.5; DB 4; Length 621;
Best Local Similarity 95.2%; Pred. No. 4.3e-225;
Matches 600; Conservative 4; Mismatches 15; Indels 11; Gaps 4;
QY 1 MALQRIDP-CWSCGDRP-GSLIFLLFSLGWHPARTLAGETGESAPLGAVLTTPHNISS 58
Db 1 MALPTARILGSCGCPALGSLIFLLFSLGWQPSRTLAGETGEAPLDGYLANPNISS 59
QY 59 LSPRLGFPCEAVESGLSTERYRELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDALPL 118
Db 60 LSPRLGFPCEAVESGLSTERYRELAVALAOKNVKLSLEQLRCLAHRLSEPPEDDALPL 119
QY 119 DLLFLFNDASSGQACRFRFSRITKANVDLLPRGAPRORLLPRAALACWGVRSLSSEA 178
Db 120 DLLFLFNDASSGQACRFRFSRITKANVDLLPRGAPRORLLPRAALACWGVRSLSSEA 179
QY 179 DVRALGLACDLPGRFVAESAELLRLVSCGPIIDODQOEARAALOGGPPYPSTW 238

```



```

Db 180 DVALGGLACLDLPGRFVAESAELLRLVSCPGPLDDQOEAAARALOGGGPPGPSTW 239
Qy 239 SVSTMDALRGILLPVLGOPITISIPQGIYAAMROSSRPSMRQPERTLLRRFRREVEXT 298
Db 240 SVSTMDALRGILLPVLGOPITISIPQGIYAAMROSSRPSMRQPERTLLRRFRREVEXT 299
Qy 299 ACPGSKKAREIDESLIFPKKWELEACVDALLATQMDVNAIPFEYEDLDVKKHLDLY 358
Db 300 ACPGSKKAREIDESLIFPKKWELEACVDALLATQMDVNAIPFEYEDLDVKKHLDLY 359
Qy 359 PGQPEVSIYHGLYFLKMSPEDIRKMNVTSLKALLEVDKGHEMSPOAPRRPLPOVA 418
Db 360 PGQPEVSIYHGLYFLKMSPEDIRKMNVTSLKALLEVDKGHEMSPOAPRRPLPOVA 411
Qy 419 TLIDRFVKGROLDKDTLDTLATFPGYLCSLSPBELSSVPPSSITWAVRPDDLDTCDDRQ 478
Db 412 TLIDRFVKGROLDKDTLDTLATFPGYLCSLSPBELSSVPPSSITWAVRPDDLDTCDDRQ 471
Qy 479 LDVLYPKARLAFQNMNGSEYFKIOSFLGAPTEBDLKAISOONVSMDLATFVKLTDAVL 538
Db 472 LDVLYPKARLAFQNMNGSEYFKIOSFLGAPTEBDLKAISOONVSMDLATFVKLTDAVL 531
Qy 539 PLTVAEVOKLLGPHVEGLKAEEHRRPVDWILRQRODDLDTLGLGLOGGIPNGYLVLDLS 598
Db 532 PLTVAEVOKLLGPHVEGLKAEEHRRPVDWILRQRODDLDTLGLGLOGGIPNGYLVLDLS 591
Qy 599 VOETLSGTPCLLGPVLTVALLLASTLA 628
Db 592 VOETLSGTPCLLGPVLTVALLLASTLA 621

```

```

RESULT 4
Q9ERA7 PRELIMINARY: PRT: 625 AA.
ID Q9ERA7
AC Q9ERA7
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE MESOTHELIN.
GN ERG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EKER RAT;
RX MEDLINE=20403886; PubMed=10944454;
RA Yamashita Y., Yokoyama M., Kobayashi E., Takai S., Hino O.;
RT "Mapping and determination of the CDNA sequence of the Erc gene
RT preferentially expressed in renal cell carcinoma in the Tsc2 gene
RT mutant (Eker) rat model."
RL Biochem. Biophys. Res. Commun. 275:134-140(2000).
DR EMBL; D87351; BAB13512.1;
SQ SEQUENCE 625 AA; 68852 MW; 343EF77ADF390CD2 CRC64;

```

Query Match 52.4%; Score 1710; DB 11; Length 625;  
 Best Local Similarity 55.9%; Pred. No. 8.2e-123;  
 Matches 352; Conservative 80; Mismatches 184; Indels 14; Gaps 5;

```

Qy 1 MALQRLDP-CMSCGDR--PGSLFLFLSLGWNVPARTLAGETGTSAPLGVLTTPHNIS 57
Db 1 MALPAPQRLDSCGSPICSRSLFLLLSLGWLPRLOQTOTRTKSQDALHNAVTV-DEFA 59
Qy 58 SLSPROLLGFCPEAVSGSLSTERVRELAVALAOKNVKLTSEDLRCLAHRISE--PPEDIDA 115
Db 60 SLPTGLTGLTCDCEVSGSLSMGHAKEELAAVAVROKNIVLQVHRCIARLRPKHLNDEIDA 119
Qy 116 LPDLTLFLNPAESGPOACTFEFSRITKANVDLPGRGAPENQRLLPALACWYRGSL 175
Db 120 LPDLTLFLNPAEPGQACAHFSLISKANVNVLPRLSREROLLTGALCKQGVYGFQV 179

```

```

Qy 176 SEADVRLGGLACLDLPGRFVAESAELLRLVSCPGPLDDQOEAAARALOGGGPPGP 235
Db 180 SETDARALGGLACLDLPGRFVAKSSVLLPMLARCGPLDDQOAKAVREVLSGARPIYGP 239
Qy 236 SWSVSTMDALRGILLPVLGOPITISIPQGIYAAMROSSRPSMRQPERTLLRRFRREV 295
Db 240 SWSVSTMDALRGILLPVLDSEIVSHIPKDVITEMLOGISREPSRLGSKMTVTHPRFRDT 299
Qy 296 EKTACPSKAREIDESLIFPKKWELEACVDALLATQMDVNAIPFEYEDLDVKKHLD 355
Db 300 EOKACPPKEENVDENLIFYONMELEACVDGTLLAGMDLVNELIPFEYEDLSIFKHLD 359
Qy 356 ELYPGPEVSIYHGLYFLKMSPEDIRKMNVTSLKALLEVDKGHEMSPOAPRRPLP 415
Db 360 KTYPGYVESLTKQGHFFRYVSPEDIRQMNVTSPDYNTILKVSKGKMD-----A 411
Qy 416 QVATLIDRFVKGROLDKDTLDTLATFPGYLCSLSPBELSSVPPSSITWAVRPDDLDTCDD 475
Db 412 QVATLIDRFVKGROLDKDTLDTLATFPGYLCSLSPBELSSVPPSSITWAVRPDDLDTCDD 471
Qy 476 PROLDVLYPKARLAFQNMNGSEYFKIOSFLGAPTEBDLKAISOONVSMDLATFVKLTDAVL 535
Db 472 ORHGLITLQKACSAFQNVSGLEYEKFIRFTFLGASREDLRALSOHNVSMDLATFVKLTDAVL 531
Qy 536 ALPLTVAEVOKLLGPHVEGLKAEEHRRPVDWILRQRODDLDTLGLGLOGGIPNGYLVLDLS 595
Db 532 ALVGLSVAEVOKLLGPHIGDGLKTEEDKSPVDWILRQROKNDLSGLGLOGGIPNGYLVLDLS 591
Qy 596 DLVSQETLSGTPCLLGPVLTVALLLASTLA 625
Db 592 DFNVEAFSSGAPLPGPVPFAMIPALLSA 621

```

```

RESULT 5
Q61468 PRELIMINARY: PRT: 625 AA.
ID Q61468
AC Q61468
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE MEGAKARYOCYTE POTENTIATING FACTOR PRECURSOR.
GN MSLN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kojima T., Taniguchi Y., Hattori K., Oh-eda M.;
RT "mouse Megakaryocyte Potentiating Factor cDNA."
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95394969; PubMed=7665620;
RA Kojima T., Oh-eda M., Hattori K., Taniguchi Y., Tamura M., Ochi N.,
RA Yamaguchi N.;
RT "Molecular cloning and expression of megakaryocyte potentiating factor
RT cDNA."
RL J. Biol. Chem. 270:21984-21990(1995).
DR EMBL; D86370; BAB13077.1;
DR MGD; MGI:1888992; Msln.
RN SIGNAL.
FT SIGNAL.
SQ SEQUENCE 625 AA; 69423 MW; 9A5E9C3222C6983F CRC64;

```

Query Match 52.3%; Score 1707; DB 11; Length 625;  
 Best Local Similarity 56.1%; Pred. No. 1.4e-122;  
 Matches 352; Conservative 80; Mismatches 182; Indels 14; Gaps 5;

```

Qy 1 MALQRLDP-CMSCGDR--PGSLFLFLSLGWNVPARTLAGETGTSAPLGVLTTPHNIS 57
Db 1 MALPAPRLDSCGSPICSRSLFLLLSLGWLPRLOQTOTRTKSQDALHNAV-NGAADFA 59

```

QY	58	SUSPOLLGFCFAEYVSGSTEREVLREVLAVLACKNVKLSSTREOLCLAHISE--PREPDLA	115
Db	60	SLPGLGFLGTCIEEVSDELSMEQAKGLANAVRKNITLRKHQRLCLARRRPHRLTDELBA	119
QY	116	LPDLLELNPDAFSPQACTRFRSRTKANVDLPRGAPERORTLPAALACMGVGSLL	175
Db	120	LPDLLELNPDAFSPQACTRFRSRTKANVDLPRGAPERORTLPAALACMGVGSLL	179
QY	176	SEADYRATLGLACDLPGRVAVASAVLLPRLVSCGPRLLDDOGEARALOGGPPYGP	235
Db	160	SEADYRATLGLACDLPGRVAVASAVLLPRLVSCGPRLLDDOGEARALOGGPPYGP	239
QY	236	STWVSSTMDALRGLLPVLGQPIRSIPQIYVAMRQSSRDSRPSWQPERTILRPREREV	295
Db	240	SKWVSSTMDALRGLLPVLGQPIRSIPQIYVAMRQSSRDSRPSWQPERTILRPREREV	299
QY	296	EKTACPSGKARDELSLFTYKKELEACVDALATOMDRVNAIPFTYEQLDVLKHKLD	355
Db	300	EKTACPSGKARDELSLFTYKKELEACVDALATOMDRVNAIPFTYEQLDVLKHKLD	359
QY	356	ELYPGYPESTVONHGYLFELKSPEDIRKMNVTSLTETKALLEYVKGHMSQAPRRPLP	415
Db	360	KTYPGYTESLLOQLGHFFRYVSPEDIRKMNVTSLTETKALLEYVKGHMSQAPRRPLP	419
QY	416	QVATLIDRFVKGRLGDKDTLDTLTAFFYVGLCSLSPHEELSVSPSSIAVAPRODITCD	475
Db	414	--IATVACYLRGGOLDDEDMVKALDIPLSTYLCDFSPQDLHSVSSVMVLVGPQDCKS	471
QY	476	PRODLVYKPAKALONNMNGSEYFVKIOSFLGCAPTEDIKALSOQVNSDLATFMRPLD	535
Db	472	ORHGLLLOKACASATONVSGLEYEFKIFLGLGASVYKDRALSOHNVSMIDATFMRPLD	531
QY	536	AVLPPLTVAEYOKLGLPHVEGLKAEERHREVRDWILRORODDLDTGLGLGGIPNGYVL	595
Db	532	SLVGSVAEYOKLGLPHVEGLKAEERHREVRDWILRORODDLDTGLGLGGIPNGYVL	591
QY	596	DLVQETLSGTPCLLGPGRVLTALLL	623
Db	592	DLVQETLSGTPCLLGPGRVLTALLL	619
RESULT	6		
Q9UK57			
ID	Q9UK57	PRELIMINARY:	PRT: 398 AA.
AC	Q9UK57		
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	MESOTHELIN/MEGAKARYOCYTE POTENTIATING FACTOR (FRAGMENT).		
DS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Scholler N., Fu N., Yang Y., Ye Z., Goodman G.E., Hellstrom K.E.,		
RT	Hellstrom I.;		
RT	"Soluble member(s) of the mesothelin/megakaryocyte potentiating factor		
RT	family are detectable in sera from patients with ovarian carcinoma.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 19:11531-11536(1999).		
DR	EMBL; AF180951; AAF01409.1;		
DR	InterPro; IPR000585; Hemopexin.		
DR	ProSite; PS00024; HEMOPEXIN; UNKNOWN_1.		
FT	NON_TER	1	
FT	NON_TER	398	398
DO	SEQUENCE	398 AA;	2D542C407807C498 CRC64;

Query Match	46.8%;	Score 1526;	DB 4;	Length 398;
Best Local Similarity	97.1%;	Pred. NO. 5.6e-109;		
Matches 297; Conservative	1;	Mismatches 0;	Indels 8;	Gaps 1;

Qy	295	VEBACPSGKKARELDESLIFKKWKELEACDALLATQMDRVAIPFTYEOQLVJKNKL	354
Db	1	VEKTAQPSGKKARELDESLIFKKWKELEACDALLATQMDRVAIPFTYEOQLVJKNKL	60
Qy	355	DELYPOGPEVSIQHLGTLFKMSPEDIRKKNVTSLETKALLEVDGHEKMSPOAPRRPL	414
Db	61	DELYPOGPEVSIQHLGTLFKMSPEDIRKKNVTSLETKALLEVKNKGHEMS-----	112
Qy	415	POVATLIDRFVKGGRQDLKDTLDTLTAFFPGYLCSSPEELSSVPSSIAWVRQDLDT	474
Db	113	POVATLIDRFVKGGRQDLKDTLDTLTAFFPGYLCSSPEELSSVPSSIAWVRQDLDT	172
Qy	475	DPROLQDLVLYPKARLAFQNMNGSEYFVKIQSFLGCAPTEDLKALSSQNVSMDLATFMKLRT	534
Db	173	DPROLQDLVLYPKARLAFQNMNGSEYFVKIQSFLGCAPTEDLKALSSQNVSMDLATFMKLRT	232
Qy	535	DAVLPVLYAEVQKLLGPRVEELKKEERKRPVRDVIILQROQDLDTLGLGIGGIGPKNGLV	594
Db	233	DAVLPVLYAEVQKLLGPRVEELKKEERKRPVRDVIILQROQDLDTLGLGIGGIGPKNGLV	292
Qy	595	LDLSVQ 600	
Db	293	LDLSVQ 298	

RESULT	7			
09F1J5		PRELIMINARY;	PRT;	698 AA.
ID	09F1J5			
AC	09F1J5;			
DT	01-MAR-2001 (TReMBLrel. 16, Created)			
DT	01-MAR-2001 (TReMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)			
DE	POTATIVE HRPO.			
GN	POTATIVE HRPO.			
OS	Pseudomonas glumae.			
OC	Bacteria; Proteobacteria; beta subdivision; Burkholderia group;			
CC	Burkholderia.			
XX	NCBI_TaxID=337;			
XX	[1]			
RM	SEQUENCE FROM N.A.			
RP				
RC	STRAIN-GM4;			
RA	Takikawa Y., Eo S., Adachi S., Kojima M.;			
RT	"Deduced hrp genes from Burkholderia glumae and Acidovorax avenae			
RT	subsp. avenae."			
RL	Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB053455; BAB20913.1;			
DR	InterPro: IPR001712; Bact_exp_FH1PEP.			
DR	Pfam: PF00771; FH1PEP.1			
DR	PRINTS: PR00849; TYPE3MAPROT.			
DR	SEQUENCE. 698 AA; 75549 MW; 4BBBE407FCE559A3 CRC64;			

	Query Match	4.1%; Score 133; DB 2; Length 698;
	Best Local Similarity	23.5%; Pred. No. 0.11; Mismatches 209; Indels 160; Gaps 31.
	Matches 135; Conservative	70; Mismatches 209; Indels 160; Gaps 31.
Oy	37 GETGTESAPLGCVLTTPHNISLSPRÖLLIGFPCAEVSGLSTERVELAVALAQKNVKST	96
	: : : : : :	:
	: : : : : :	:
Db	218 GMSADADANREFSVSLVGDAWVSQIPSLT---	SYAAGVMITRVADEROA--KORSIGD 270
Oy	97 EOLRLARLRSPEDDLALPLDLLF-LNPD-----AFSG----	POAC 135
	: : : : : :	:
	: : : : : :	:
Db	271 E----IGNQLOSSSALFFAAVILLALFAIVRGFPBSLLEVLAAALSFGYRLSYKKPSSR	326
	: : : : : :	:
	: : : : : :	:
Oy	136 TRFFSRITKANVDLLPGCAPERÖRLPPAA--LAC-WGYRGSLSEADYRAIIGLACDLPG	192
	: : : : : :	:
	: : : : : :	:
Db	327 DRAMRETLASMQRYGR-KTEVPILLPAPAFACAGVR-----ISPDL--	369
Oy	193 RFVAESAEEVLLPRVSCGCPDDDOGEARALAQG-GPPYGPPSTW-----SVSTMAL	246
	: : : : : :	:
	: : : : : :	:
Db	370 -----AAAVNAFTLDMA-----FEARALALQALGLPFPGITMWWHAPLVATIFELL	416
Oy	247 RGLPEVL-----GO---PIIRSIPOGIIVAAWRÖRSSRDPSMRÖPERTILRRFRREVEK	297

```

Db 417 IHDPRLHMTWMPKAMLPDLRLVPAEYAAA-----SARSGWMPRR-----RGRPPVDS 466
QY 298 TACS-GKKAREIDESLIFFKKWLE-----ACVDALLATQMDRVNAIFPTTEQL 347
Db 467 AGFRLRLRRAGACODAV-----NRTEQIIAHLVAPCGATFLFGLGIEYOMI-----L 516
QY 348 DVLNKKHKLDELPGYRPESVIOHLYGLFKMSPEDIRKMNWTS-LETLKALLEVKGHEMS 406
Db 517 DQSHDAPGLVAEYQKVLPPORADVLRLEEOVSIRNWSFMSLWAGAKERDMLML 576
QY 407 POARRRPLPOVATLIDREVKRGOLDKDTLDTLTAFFPGYLSLSPPE-----LSVPPS 461
Db 577 TEYVRGDSRF--LAHNAKGERQLS-----AVLFQMAQDHIAAKIKQPTG 622
QY 462 SIWAVRPODLTCPPROLDVLYPKARLAFONMNGSEYFVKIQSELGAPTEDKALSQON 521
Db 623 NFLALPMD-----AAF-----LIDRIQSEVGAAPREGVVLVT--- 655
QY 522 VSMDLATEFKLRTDAVLP-LTFAEYOKLGPHE 554
Db 656 -SMDIRRVRRMISRLGMLSVYSIQE-LGEHVE 687

```

```

RESULT 8
Q9UGS3 PRELIMINARY: PRT: 797 AA.
ID 09UGS3
AC 09UGS3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE D4756623.1 (NOVEL LEUCINE RICH PROTEIN) (FRAGMENT).
GN D4756623.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Corby N.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035681; CAB63072.1;
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003591; LRR_typ.
DR InterPro: IPR003592; LRR_out.
DR Pfam; PF00560; LRR; 18.
DR Pfam; PF01462; LRRNT; 1.
DR Pfam; PF01463; LRRCT; 2.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00370; LRR; 8.
DR SMART; SM00082; LRRCT; 2.
DR SMART; SM00013; LRRNT; 2.
DR SMART; SM00369; LRR_Typ; 4.
FT NON_TER 797
FT SEQUENCE 797 AA; 85614 MW; 8C3247883EAE59AD CRC64;

```

```

Query Match 4.0%; Score 131.5; DB 4; Length 797;
Best Local Similarity 24.4%; Pred. No. 0.17;
Matches 113; Conservative 45; Mismatches 181; Indels 125; Gaps 23;
QY 15 RPSGL--LFLFSGWHPARTLAGETESAPLGGLVITTHRN-ISSLSPQOLGFPQAE 71
Db 191 RRGFGALGALATLNLAINALVYLPAMAFQGLLRVRLRLSHNMLSVLAPALAGLPALR 250
QY 72 VSGSTEREVELAVALAOKNVKLTSTEOLRCLAH-RLSEPP-----EDLDLPL--LDLTL 122
Db 251 RLSLHNHLQALPGPV-----LSQARGLARLELGNPLTLTYAGEEGCLALPGIRELL 302
QY 123 -----FLNPDAFSGPOACTRFFSRITKAN--VDLLP-----RGAPERQRLPALAAC----- 167

```

```

Db 303 DGGALQALGPRAFA---HCPRLHTLDRGNQDITLPLPGGPGQLRLRLQGNPLMCCQA 359
QY 168 -----WVGRLSLSEADYRALGGLACDLPGRFVAESAENVLLPRLVSCPGPLDQOQEAR 222
Db 360 RPLEW-----LARARVSDG--ACQPRRLRGEALDALRPMQLRCPGDAAOEELELE 411
QY 223 AALQG-----GPPYGPSTWSVTMDALRGLLPVLGPI-----ISIPQIVAAWROR 272
Db 412 RAVAGPRAPRPPGPRGGEERAVALPCPRACVCYVESHSSEGGGLQAVPRGF----- 464
QY 273 SSRPSWRQPERTILPRFRREVEKTACPSGKKAREIDESLIFFKKWLEACVDALLAT 332
Db 465 -----PS--DTQLDLRLRHNPSPVRAAF-----GLGHLVSLHLCGIAELEAG 508
QY 333 QMDRVNAIFPTY---EQLDVKKHKLDELXPGYRPESVIOHLYGLFKMSPEDIRKMNWTS 389
Db 509 ALAGLGRILTYLSDNQLAGLSAALLEGAPR-----LGYLYLEKN----- 548
QY 390 LETLKALLEVDKGHEMSPOARRRPLPOVATL-----IDREVKG 427
Db 549 -----RFLQV-----PGALRALPLSLHLQDNNAVDRILAPG 580

```

```

RESULT 9
Q9WXJ0 PRELIMINARY: PRT: 939 AA.
ID 09WXJ0
AC 09WXJ0;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ACONITASE (FRAGMENT).
GN ACON.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamura J., Kimura E., Oosumi T., Nakatsu T.;
RT "Brevibacterium lacticolementum ATCC 13869 acn gene for Aconitase.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB025424; BAA76717.1;
DR HSSP; P16276; 180J.
DR InterPro: IPR001030; Aconitase.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam; PF00330; aconitase; 2.
DR Pfam; PF00694; Aconitase_C; 1.
DR PRINTS; PR00415; ACONITASE.
DR PRODOM; PD000511; Aconitase.
DR PROSITE; PS00450; ACONITASE_1; UNKNOWN_1.
DR PROSITE; PS01244; ACONITASE_2; 1.
FT NON_TER 1
FT SEQUENCE 939 AA; 101771 MW; E2DE6472D7F4A096 CRC64;

```

```

Query Match 3.9%; Score 126; DB 2; Length 939;
Best Local Similarity 20.3%; Pred. No. 0.56;
Matches 137; Conservative 76; Mismatches 205; Indels 258; Gaps 35;
QY 60 SPROLL-----GPPCAEVSGLSTERVELAVALAOKNVKLTSTEOLRCLAHRLSEPPEDLD 114
Db 79 TPARYLMDFTGVCPC--VVDLAT--MRAVVA-----LGQDPNDVN 115
QY 115 AL-PLDLL--FLNPDAFSGPOACTRFFSRITKANVDLLPRGAPERQRLPALAACGVR 171
Db 116 PLNPAEVYIDHSYIVEAFGRPDALAK-----NVEIEYRNEERYQFLR-----WGSE 162
QY 172 GSLSSEADYRALGGLACDLPGRFVAESAENVLLPRLVSCPGPLDQOQEAALQGGPP 231
Db 163 -----SFSNFRVVP-----PQTGIVHOVNI-----ETLARVVEPDNECLA 196
QY 232 YGPSTW-----SVSTMDALRGLL-----PVLGQPIIRSIPOGIIVAAWRQSSRD 276

```





RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.,  
 RT "Sequence and analysis of chromosome II of *Arabidopsis thaliana*."  
 RL Nature 402:761-768(1999).  
 DR EMBL: AC006202; AAC29825.1;  
 SO SEQUENCE 2218 AA; 235738 MW; 525B6B1937A62998 CRC64;

Query Match 3.7%; Score 120; DB 10; Length 2218;  
 Best Local Similarity 19.1%; Pred. No. 5.7;  
 Matches 135; Conservative 81; Mismatches 222; Indels 270; Gaps 32;

QY 30 HPARTLA-----GETGESAPLGGVLTTPHNISLSPPROLLGFPACAEVSGLSTERYRE 82  
 DB 169 HPTSSALTSFDLSGPPFOSLP-----ASPAPTPIRGGRSGRSGRGAAGRR 217  
 QY 83 LAVALAOKNVKLS--TEQRLCIAHRLSEPPEDLALPDLILLFLNPDASFSGPOACTRFFS 140  
 DB 218 VEGVLHGSNNSITORTETATSLA-----SDAEA-----TFALPRASSELVS 259  
 QY 141 RITRANVDLPRGAPER-ORLLPALACGWYRGSILSEADYRVALGGLCDLPGRFAESA 199  
 DB 260 RVPKAN-----EGSTSNPDQVSPVHSATTALRSOKAKDKL-----DAPGFGSGSH 306  
 QY 200 EVLLPRLVSCGPIIDQOGEAA-----RAALQGGGPPYGPSTMSV-----STMD 244  
 DB 307 -----VQTLNVLNLSSEKRAFAVKRRPLIOGGGPVONONAVSSVCGSGSPSEGRITYT 359  
 QY 245 ALRGL-----LPVLGQPI-----IRSIPOGIYAAMQR--- 272  
 DB 360 ALQGVTPAPSAITLPMSSQPSDAILPMSSQVGTVEAOEANVPSLPALPAKRRVRLP 419  
 QY 273 SSRPSWRQPER-----TIL-----RPRFR 293  
 DB 420 SRGETPKRQGRRGQPLPATDASSARSTGLTPQLEVKVGNLSGTAKPKDAYAKEQPHFSQ 479  
 QY 294 EYERTACSGSKKAREIDSLIFKKWELEACVDALLATQMDRVNAIFPTTEQLDVLYLKH 353  
 DB 480 SVAPDIHSSGSLSOEIRDTSGTGSARKQATDVTVARVKEI-----FSETSLKHK 533  
 QY 354 LDELPGQVPSVIOHGLFLKMSPEDIRKMNVTSLTLK-----ALLEVKG 402  
 DB 534 VGE--PSATTTTNVPD-----AOSPGEM--NLHTVETHKAEDSSGLKNOBALYNLSKA 582  
 QY 403 HEMSPQARRRPLPOVATL-----TAFYGYLCSLSPELSV-----IDRF 424  
 DB 583 DKIVSDIP-HVPGLDTTSGSVANKVDIGSSKYAAENELVKIPGVDSSVIOISLNT 641  
 QY 425 VKGRQLDKDTLDL-----TAFYGYLCSLSPELSV-----PPSITMAVR 467  
 DB 642 LTAKSSLEKCTADOLLEKLSQEGETTPASDGETCHLAETFASSISYVRSPTASA--- 697  
 QY 468 PODDTCOPRQDLVLYPKARLAFON---MNGSEFYVKIQSEFLGAPREDIKALSOQVNS 523  
 DB 698 ---STTAEPPLPTDKL--EKNISFODEVKTUNG---KREAIL-----LSSEQTIVN 741  
 QY 524 MDLAT---FMKLRDVAVLPTVAEVOKLGPHEGLKAERHRVRD 567  
 DB 742 SKIETNSELOASRTDEV-----PHVDGKSVDVANQTVKE 776

RESULT 14  
 ID 09AUB4 PRELIMINARY; PRT; 3574 AA.  
 AC 09AUB4;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DE PUTATIVE CHROMATIN REMODELING PROTEIN SYD.  
 GN SPLAYED.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCB1\_TaxID=3702;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Wagner D., Meyerowitz E.M.;  
 RT "SPLAYED, a Putative Chromatin Remodeling Factor, Controls  
 RT Reproductive Development in Arabidopsis."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF247809; AAK31908.1;  
 SO SEQUENCE 3574 AA; 389826 MW; 0CA25C1F1AB6A0 CRC64;

Query Match 3.6%; Score 117.5; DB 10; Length 3574;  
 Best Local Similarity 19.2%; Pred. No. 18;  
 Matches 138; Conservative 82; Mismatches 233; Indels 265; Gaps 32;

QY 30 HPARTLA-----GETGESAPLGGVLTTPHNISLSPPROLLGFPACAEVSGLSTERYRE 82  
 DB 1500 HPTSSALTSFDLSGPPFOSLP-----ASPAPTPIRGGRSGRSGRGAAGRR 1548  
 QY 83 LAVALAOKNVKLS--TEQRLCIAHRLSEPPEDLALPDLILLFLNPDASFSGPOACTRFFS 140  
 DB 1549 VEGVLHGSNNSITORTETATSLA-----SDAEA-----TFALPRASSELVS 1590  
 QY 141 RITRANVDLPRGAPER-ORLLPALACGWYRGSILSEAD-----VRALGGLA 187  
 DB 1591 RVPKAN-----EGSTSNPDQVSPVHSATTALRSOKAKDKLDAPGFGSGSHVOTLNVLE 1645  
 QY 188 CDLPGRFAESADEVLLPRLVS--CGEPDLO--DOGEAANAALQGGGPPYGPSTMSV--- 240  
 DB 1646 NSSERKARFAVKRRPLIOGVSSOHGPNKOPDLVPVSTSLTGGGPVONONAVSSVCGS 1705  
 QY 241 ---STMDALRGL-----LPVLGQPI-----IRSIPOGI 265  
 DB 1706 KSPSGRITYTLQGVTPASDAILPMSSQVGTVEAOEANVPSLPAL 1765  
 QY 266 VAAMRQ---SSRPSWRQPER-----TIL----- 287  
 DB 1766 PAKRRVRLPNSGETPKRQGRRGQPLPATDASSARSTGLTPQLEVKVGNLSGTAKPKDA 1825  
 QY 288 ---RPRFRREYERKACSGSKKAREIDSLIFKKWELEACVDALLATQMDRVNAIFPT 343  
 DB 1826 VAKQEPHFSSQVAPDIHSSGSLSOEIRDTSGTGSARKQATDVTVARVKEI----- 1879  
 QY 344 YEQDLVLYLKHDELTPQGVPSVIOHGLFLKMSPEDIRKMNVTSLTLK----- 394  
 DB 1880 FSEISLKHKGE--PSATTTTNVPD-----AOSPGEM--NLHTVETHKAEDSSGLKN 1928  
 QY 395 --ALLEVDKHEMSQARRRPLPOVATL----- 420  
 DB 1929 QEALYNLKSKADKIVSDIP-HVPGLDTTSGSVANKVDIGSSKYAAENELVKIPGVDVS 1987  
 QY 421 ---IDRFKGRQLDKDTLDL-----TAFYGYLCSLSPELSV----- 458  
 DB 1988 SVIOLSLNTLTAKSSLEKCTADOLLEKLSQEGETTPASDGETCHLAETFASSISYVRS 2047  
 QY 459 -PPSITMAVRQDDTCOPRQDLVLYPKARLAFON---MNGSEFYVKIQSEFLGAPTED 513  
 DB 2048 EPTASA-----STTAEPPLPTDKL--EKNISFODEVKTUNG---KREAIL----- 2087  
 QY 514 LKALSOQVNSMDLAT---FMKLRDVAVLPTVAEVOKLGPHEGLKAERHRVRD 567  
 DB 2088 LSSEQTIVNNSKIETNSELOASRTDEV-----PHVDGKSVDVANQTVKE 2132

RESULT 15  
 ID 09U5Y1 PRELIMINARY; PRT; 2015 AA.  
 AC 09U5Y1;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-JUN-2001 (TREMblrel. 13, Last sequence update)  
 DE MICROTUBULE-ASSOCIATED PROTEIN CP224.

OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX2;  
RA Graef R., Daubderer C., Schliwa M.;  
RT "Dictyostelium DdcP224 is a microtubule-associated protein and a  
permanent centrosomal resident involved in centrosome duplication.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ012088; CAB56504.1;  
DR InterPro: IPR000357; HEAT\_repeat.  
DR PROSITE: PS50077; HEAT\_REPEAT; 1.  
SO SEQUENCE 2015 AA; 224038 MW; BA64E982ADDC92EE CRC64;

Query Match 3.6%; Score 116.5; DB 5; Length 2015;  
Best Local Similarity 23.3%; Pred. No. 9.2;  
Matches 66; Conservative 30; Mismatches 74; Indels 113; Gaps 13;

QY 355 DELYPGYPESVIOHLYFLIKMSPEDIRKKNVTSLET----- 393  
Db 3 DEAPSGSIEDRIHKKM-----KRVSGLEELTTKFRNSIEGSGPLFNEWGPQ 51  
QY 394 -KALLEVDKGHEMSPOAPRRPLPOVATLIDRFVKGRLDKDTLDTLTAFFPYGL----- 447  
Db 52 FKKILA-----DINPMGOERALEPLSAFIDRC-----DCVNKFAASYVGVLEKLF 97  
QY 448 -----CSLSPBELSSVPP-----SSIWAVRPDLDTCDP 476  
Db 98 ASTRPRAKEKTIIECLLTLEADSAPVVEALLKGTSTSPKILLASL-AALTOALKTFGP 156  
QY 477 ROLDVLPKARLA--FONMN-----GSEYFVKIOSFLGA-----PTEDLKALS 518  
Db 157 KQIVYKLLIKQFSPWFENRDKGIRDQASSELFIEIYRWIGKALILPILISEALTPIQ-LKALQ 215  
QY 519 QQNVSMDLATFMKLRTPDAVLPL--TVAEVOKLLGPHVEGLKAE 559  
Db 216 DQ-----FEKLPPTPAVPLKTYRSEAAKALANASKGIOAK 250

Search completed: December 7, 2001, 09:53:43  
Job time: 326 sec

**THIS PAGE BLANK (USPTO)**